

**LOUGHBOROUGH
UNIVERSITY OF TECHNOLOGY
LIBRARY**

AUTHOR/FILING TITLE

FIADDEIRO, P.T.

ACCESSION/COPY NO.

040121822

VOL. NO.

CLASS MARK

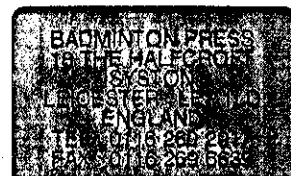
LOAN COPY

27 JUN 1997

26 JUN 1998

23 MAY 1999

0401218228



WAVE-FRONT RECONSTRUCTION OF OPTICAL DISTURBANCES USING DIGITAL IMAGE PROCESSING

by

Paulo Torrão Fiadeiro

A Doctoral Thesis

Submitted in partial fulfilment of the requirements for the award of Doctor of
Philosophy at the Loughborough University of Technology.

December 1995

© by P. T. Fiadeiro 1995

Loughborough University Office of the Librarian	
Date	May 96
Class	
Acc. No.	640181822

9 6446783

ACKNOWLEDGEMENTS

I would like to take this opportunity to thank my parents, family and friends from whom I have received warm moral support during the period of my research. I would especially like to mention my supervisor Dr. David Emmony who has provided many interesting discussions and much guidance throughout the past three years.

I would also like to acknowledge the generosity of the Calouste Gulbenkian Foundation, Lisbon (Portugal), for the financial support of the work in this thesis under the grant ref. 8/92/B. I am indebted to Prof. Carlos Sá Furtado, the director of the Scholarship Office, and to Mrs. Maria Antónia da Silva, the director of the Portuguese Affairs Department, from whom I have received a great help and who have made this thesis possible.

An additional thank you to all those people at Loughborough and at Covilhã (University of Beira Interior, Portugal) who have given their assistance and encouragement. My thanks also go to John Oakley, Ramesh Pancholi, and Mahendra Pancholi in the Physics Department Workshop and to all the research students in the Optoelectronics Laboratory.

To my Parents and in memory of my nephew José António
(Para os meus Pais e em memória do meu sobrinho José António)

ABSTRACT

This thesis is concerned with the development of a practical digital image processing system for recording and subsequent reconstruction of the magnitude and phase of an optical wave-front arriving from a coherently illuminated object disturbance.

Since the wave-fronts of concern are coherent, the magnitude and phase of such waves are generally independent functions in the sense that the knowledge of one is not sufficient to uniquely deduce the other. To uniquely reconstruct and characterize optical disturbances both the magnitude and phase are required. In general, all recording media respond only to light intensity and no difficulty is encountered in recording the intensity and therefore the magnitude, because it is the square root of the intensity. It is therefore necessary that the phase information be somehow converted to light intensity variations for recording purposes. Several techniques, such as interferometric techniques, are normally used to convert phase information into light changes. The practical implementations of such techniques are not always simple. An alternative to these techniques, based on digital Fourier transformations, is proposed and applied to the phase retrieval problem.

Computer simulations and practical results show that the Gerchberg and Saxton iterative algorithm is suitable for recovering the phase information from single measurements of the intensity in both the image and Fourier domains in an optical system. A considerable improvement in the experimental results is obtained when non-linear image sensors are used to perform quasi-magnitude measurements instead of the usual intensity measurements.

The magnitude and phase of the digitally reconstructed wave-front are used to characterize the corresponding optical disturbance. In general, optical disturbances are created in transparent media and they are characterized by refractive index changes that deviate the incident light from its normal trajectory and consequently the phase of the wave-front is changed. Knowing the phase it is possible to perform measurements of the refractive index, density, and some other parameters of the medium under study. The generation of schlieren images on the computer screen is also possible and the method proved to be a good complement to the conventional schlieren technique due to the high sensitivity that is normally not achievable in practice.

CONTENTS

INTRODUCTION	1
 CHAPTER 1: Characterization Techniques of Optical Disturbances	
1.1 Introduction	4
1.2 Shadowgraphy	6
1.3 The Schlieren Technique	8
1.4 Interferometric Techniques	13
1.5 An Alternative to Interferometric Techniques	17
1.5.1 Historical Review	18
1.5.2 The Proposed Technique	21
1.6 Summary	22
References	23
 CHAPTER 2: A Digital Image Processing System	
2.1 Introduction	28
2.2 Digital Image Fundamentals	29
2.2.1 The Image Model	30
2.2.2 Sampling and Quantification	31
2.3 Elements of a Digital Image Processing System	33
2.3.1 Image Detectors	34
2.3.2 Image Processors	37
2.3.3 Digital Computers	40
2.3.4 Storage Devices	41
2.3.5 Display and Recording Devices	41
2.4 The Developed System	42
2.4.1 Hardware Description	42
2.4.2 Software Description	44
2.4.3 External Trigger and Synchronization	46
2.5 Summary	47
References	48

CHAPTER 3: Image Transformations

3.1	Introduction	50
3.2	Continuous Fourier Transformation	50
3.3	Discrete Fourier Transformation	53
3.4	Fourier Transform Properties	58
3.4.1	Separability	59
3.4.2	Translation	60
3.4.3	Periodicity and Conjugate Symmetry	61
3.4.4	Rotation	62
3.4.5	Distributivity and Scaling	62
3.4.6	Average Value	63
3.4.7	Parseval-Rayleigh Theorem	63
3.4.8	Laplacian	64
3.4.9	Convolution and Correlation	65
3.4.10	Sampling Theorem	70
3.5	Fast Fourier Transform	76
3.5.1	FFT Algorithm	78
3.5.2	Inverse FFT	79
3.6	Summary	80
	References	82

CHAPTER 4: Image Formation and Reconstruction

4.1	Interference and Diffraction	84
4.1.1	Huygens' Principle	85
4.1.2	Young's Experiment	86
4.1.3	Fresnel's Biprism	88
4.2	Fresnel and Fraunhofer Diffraction	90
4.2.1	Diffraction by a Single Slit	91
4.2.2	Diffraction by a Rectangular Aperture	96
4.2.3	Diffraction by a Circular Aperture	98
4.3	Image Formation	101
4.4	Image Reconstruction	107
4.4.1	The Importance of Phase	109
4.4.2	The Gerchberg and Saxton Algorithm	112
4.5	Summary	118
	References	120

CHAPTER 5: Phase Retrieval for Wave-Front Reconstruction

5.1	Introduction	124
5.2	Methodology	124
5.3	Computer Simulations	126
5.3.1	Intensity and Magnitude Measurements	127
5.3.2	Simulation Results	129
5.3.3	Conclusions	142
5.4	Experimentation	143
5.4.1	Experimental Arrangement	143
5.4.2	Set-up, Scaling and Sampling	144
5.4.3	Intensity and Quasi Magnitude Measurements of Light	153
5.4.4	Experimental Results	157
5.4.5	Conclusions	171
5.5	Summary	172
	References	174

CHAPTER 6: Computer Generation of Schlieren Images

6.1	Introduction	175
6.2	Wave-Front Reconstruction and Schlieren Images	176
6.2.1	Simulated Example	177
6.2.2	Real Examples	184
6.3	Conclusions	191
	References	193

CHAPTER 7: Conclusions and Further Work 194**PUBLICATIONS: Articles and Conference Papers** 197

INTRODUCTION

The dominant characteristic of the *technology age* is perhaps the multitude of measurement apparatus, diagnostic devices, and data transmission and reception systems which have spread over most facets of human existence nowadays. Much of the data gathered by such systems can be regarded as possessing intensities or magnitudes and phases. The relative accuracy, with which intensity and phase can be sensed, depends upon the range of frequencies, or equivalently wavelengths, of the signals carrying the information.

For frequencies of a few MHz or less, phase can often be measured more precisely than intensity. For higher frequencies, however, the accurate determination of phase becomes progressively more difficult. When one reaches the optical bands, infra-red and visible light, phase can no longer be directly recorded. It has, in general, to be inferred from interference patterns, as in holography and its adaptations to quantitative sensing of surface shape and displacement. In the ultra-violet and beyond, phase can only be recovered by performing lengthy digital computations, because it is virtually impossible to observe the kinds of interference from which phase can be deduced.

During the past three years the particular problem of recording and subsequent reconstruction of the magnitude and phase of an optical wave-front arriving from a coherently illuminated (visible light) object disturbance, without using interferometric techniques, has been studied. This problem is sufficiently general to be of interest for electromagnetic waves, as well as for acoustic waves. However, the case discussed here will be restricted to the optical problem. This thesis describes work which has been carried out in that time, from the development of a digital image processing system, which has been used as a simulation tool, to the practical application of such a system to real experiments for optical wave-front reconstruction.

Chapter one is an introduction and review entitled *Characterization Techniques of Optical Disturbances*. Several techniques such as shadowgraphy, schlieren and interferometry, that are normally used for qualitative and quantitative characterization of optical disturbances, are described. Finally, an alternative to interferometric techniques,

based on Fourier transforms, is discussed and proposed for optical wave-front characterization.

Chapter two, *A Digital Image processing System*, gives the digital image fundamentals and describes the elements in a basic general purpose digital image processing system. Hardware and software considerations concerning the system that have been developed for digital image processing are also presented in detail.

Chapter three, *Image Transformations*, presents the theory concepts and properties of the Fourier transformation which provided the practical implementation of the proposed technique for the characterization of optical disturbances using digital image processing. An algorithm for the fast Fourier transform computation is also fully described.

Chapter four, *Image Formation and Reconstruction*, describes the optical image formation and reconstruction processes under coherent illumination. A brief review of interference and diffraction, illustrated with several examples, as an intermediate step on the image formation process, and the importance of the phase in image reconstruction are given. A complete description of the operation of the implemented algorithm, based on the Gerchberg and Saxton algorithm, for phase retrieval using Fourier transforms, is also presented.

Chapter five, *Phase Retrieval for Wave-Front Reconstruction*, describes the methodology of application the algorithm for phase retrieval information and also addresses its practical problems and limitations. Computer simulations that have been carried out to recover phase information using the implemented algorithm are presented and described. Finally, the description of practical real experiments and realizations which have been adopted to apply the technique for phase retrieval and the corresponding results have also been included.

Chapter six, *Computer Generation of Schlieren Images*, describes a technique that complements the conventional schlieren technique by generating and analysing schlieren images on the computer screen. This technique is basically a frequency domain manipulation of the data in the reconstructed wave-front. The results of simulated and real examples which have been carried out are also presented and described.

Chapter seven, *Conclusions and Further Work*, summarizes the conclusions which have been made during the course of this work and suggests possible areas and applications for further work.

Finally, the section entitled *Publications* presents a list of articles and paper conferences which have been published and presented by the author over the period of study.

CHAPTER 1

Characterization Techniques of Optical Disturbances

1.1 Introduction

Many problems of science and engineering involve substances that are colourless, transparent and non-luminous, so that their observation by direct visual or photographic methods is difficult. Optical disturbances, also called optical non-uniformities or non-homogeneities, are created in transparent media and they are characterized by a variable refractive index. The purpose of investigating optical disturbances lies in determining the refractive index in different parts of a given medium with the subsequent calculation of densities, pressure, temperature and other parameters in order to characterize the state of the substance.

According to the studies conducted by Lorentz [1], the relationship between the density, ρ , of a transparent substance and its refractive index, n , can be presented as follows:

$$\frac{n^2 + 1}{n^2 + 2} = K\rho, \quad (1.1-1)$$

where K is a constant factor for the given substance. For a gaseous medium, the relation between ρ and n is called the Gladstone-Dale ratio [2] and has the following form

$$n - 1 = K\rho. \quad (1.1-2)$$

If the values ρ_0 and n_0 are known at certain standard temperature and pressure, the previous relation, defined in Eq. (1.1-2), can be transformed and rewritten as

$$\frac{n - 1}{\rho} = \frac{n_0 - 1}{\rho_0}. \quad (1.1-3)$$

Thus,

$$\rho = \frac{n - 1}{n_0 - 1} \rho_0 \quad (1.1-4)$$

which means that knowing the distribution of the refractive index for the given medium and for a certain standard medium, one can transfer to density distribution in the disturbed region.

Shadow, schlieren and interference methods of research are used for determining changes of the refractive indices of transparent media. These methods are widely applied in the study of perturbations arising in the air, in a gaseous medium, or in a liquid, and also in transparent solid substances. Changes of the refractive index are converted into changes of the illumination on a screen or photographic film by means of special equipment. This method is called optical visualization.

To describe the principle underlying this method of optical visualization, suppose a ray of light enters a certain disturbed medium Q , as shown in Fig. 1.1. On passing through this medium, the ray falls upon a screen or a photographic film P , but at the same time it deviates from its initial direction. Two parameters can define the position of the ray after it passes through the disturbed medium. The ray does not arrive at point x , where it would have been on passing a homogeneous medium, but at point x' . The displacement between x and x' can be measured. The ray is deflected by an angle φ which can also be measured. Moreover, since the speed of the light changes depending on the density of the medium [3], the ray reaches points x and x' at instants of time t and t' , respectively, which can also be measured.

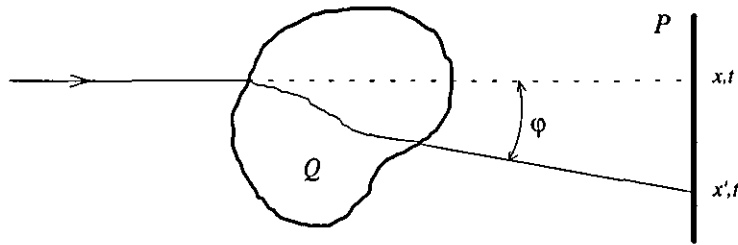


Fig. 1.1 - Rays of light passing through a disturbed or non-uniform medium.

These three quantities form the basis for three methods of optical visualization of disturbances [4], namely, *shadowgraphy* which defines the linear displacement of the ray of light, the *schlieren technique*, which defines the angular deflection of the ray, and the *interferometric technique*, which defines the difference in time. These techniques of research make it possible to make a qualitative and quantitative assessment of the medium being studied, of the shape and propagation velocity of waves, of the distribution of density, pressure, temperature, etc. The principal merits of these techniques are:

- light does not affect the area being studied;
- it is possible to make measurements over the entire field simultaneously.

Since many of the events studied by these techniques develop very rapidly, it is also possible to record the changes of the refractive index by using high-speed photographic

cameras. For this purpose the cameras should be arranged in a certain ways with shadow, schlieren and interferometric instruments.

1.2 Shadowgraphy

The term shadowgraphy refers to the method developed by Dvorak [5], in which an optical disturbance is illuminated by a point light source, and some section of the light beam beyond the non-homogeneity is observed.

Figure 1.2 presents the schematic diagram of an instrument for shadowgraphy. A ray of light from the point light source S passes through the field being studied Q . Wherever this field is disturbed or optically non-uniform, the ray is deflected from its initial direction and a shadow pattern is observed on the screen P .

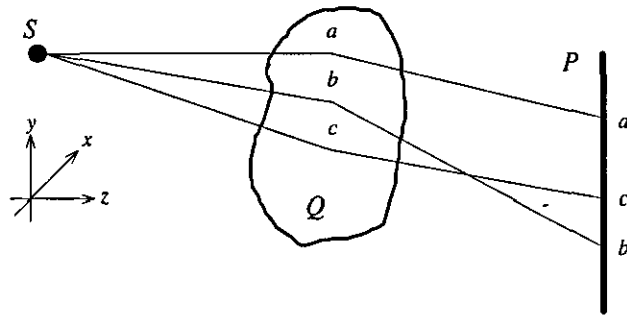


Fig. 1.2 - Schematic diagram illustrating the shadowgraphy method.

If the gradients of the refractive index $\partial n/\partial x$ and $\partial n/\partial y$ throughout the length of the area under investigation remain constant, all the rays will be deflected equally so that no change of illumination is observed on the screen. If $\partial n/\partial x$ and $\partial n/\partial y$ changes, that is, the second derivatives $\partial^2 n/\partial x^2$ and $\partial^2 n/\partial y^2$ are not equal to zero, the deflection of the rays will vary. In Fig. 1.2, three points a , b , and c are shown in the area under investigation Q . If the gradient at point b is larger than at points a and c , the rays passing through the area of point b will deflect more and the track of the ray on the screen coming from point b , namely, point b' will be outside of the segment $a'c'$. A shadowed area will appear between a' and c' , whereas the area between c' and b' will be lighter.

Figure 1.2 shows a diagram where the distance from the source of light S to the medium studied Q is assumed to be very large (several meters) since the source of light usually has finite dimensions. When using a source of light which is practically a point source, the given distance may be considerably reduced. The presented layout, however,

produces a shadow pattern that is not focused on the screen P , and this reduces the clarity of the created shadow images. If an optical system L is placed between the region under investigation Q and the screen P , one can produce a focused shadow pattern on the screen when the cross section abc and the plane of the screen lie in optically conjugate planes, as shown in Fig. 1.3. The three points a , b and c are imaged respectively to a' , b' and c' on the screen P through the optical system L . However, the light intensity distribution on the screen will appear with a better contrast between the lighter and dark areas because some of the deflected light by the medium Q is not collected by the finite aperture of the system lens. If the optical system has an infinite aperture all the light from the field will be collected and no effect on the image contrast will be observed.

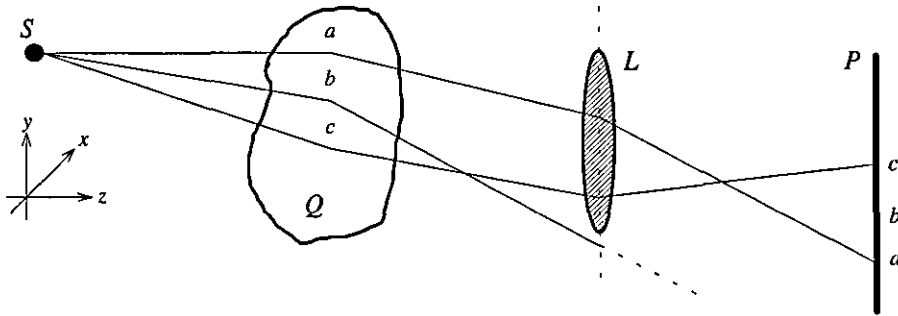


Fig. 1.3 - Schematic diagram of the instrument with supplementary lens.

This focusing procedure considerably increases the contrast of the shadow image. A sharper pattern may also be produced if collimated rays of light are used as the source of illumination. As an example, Fig. 1.4 shows a shadow image of a shock wave propagating in water. Two adjacent stripes, one light, the other dark, are easily observed because the pressure increases sharply at the front of the wave and then gradually decreases in the direction of the source of the disturbance.

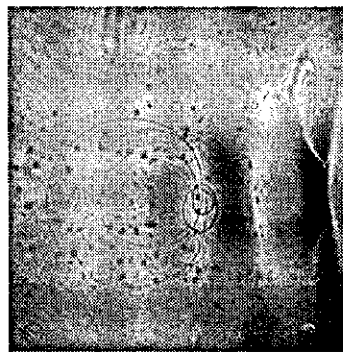


Fig. 1.4 - Shadow image of a shock wave in water.

The shadowgraphy technique [6,7] makes it possible to study the changes of the refractive index not only in one given direction, but in the entire non-uniform medium

or disturbance so that in the general case it can be stated that the change of illumination of the shadow pattern is subject to the law:

$$\Delta I = K \int \left(\frac{\partial^2 n}{\partial x^2} + \frac{\partial^2 n}{\partial y^2} \right) dz, \quad (1.2-1)$$

where K is a constant, x and y are co-ordinates in a plane normal to the ray of light, and z is the co-ordinate in the direction of the ray. The second derivative of the refractive index can also be determined by shadowgraphy because the effects observed with this method are a function of that derivative as expressed by Eq. (1.2-1). It is sensitive only to sharp changes of the refractive index and therefore often applied in investigating phenomena with a sharply defined front. Double integration makes it possible to convert immediately to the value of the refractive index of the given area of the region under investigation.

Shadowgraphy is insufficiently sensitive for the study of slow and smooth changes in the density of a medium. For this reason other methods or techniques should be used to cover these cases.

1.3 The Schlieren Technique

The schlieren technique was suggested by the French astronomer Foucault [8,9] in 1858 as a means of quality control in the manufacture of large astronomic objectives of high resolving power. In this sense it is known to this day as the Foucault knife-edge test. In 1864, the German physicist, Töpler, applied this method to the study of gaseous inhomogeneities [10]. In effect, he uncovered a new application of the already known Foucault method, but the great value of the results obtained and the increasingly large role played by gas dynamics in the general development of science have served to attach Töpler's name to this method. Up to the 20th century, the optical quality tests and the study of gaseous non-homogeneities were the only areas for the application of the schlieren technique. With the advent of the 20th century, schlieren methods underwent considerable development, being possible to establish a transition from qualitative to quantitative measurements. These methods acquired a physical interpretation, based on the laws of optical waves [11,12], and their importance has increased in recent years, in connection with the development of the diffraction theory [13].

The basic idea behind the schlieren technique is to detect lateral displacements of rays (wave-fronts may be deformed) by blocking out or modifying these displaced rays. The

blocking or modification can be accomplished by placing screens in any of the planes of convergence of the light passing through, or being reflected from, the optical region under test. Figure 1.5 shows the principle of the schlieren method to perform qualitative measurements. It consists of the following: a point light source S is set at some distance from an objective L , Fig. 1.5(a), or from a spherical mirror M , Fig. 1.5(b); a plate with a sharp straight edge K , called the Foucault knife-edge, is placed in the image plane of the optical system. If the objective is free of aberrations and the mirror is perfectly spherical, the transmitted and reflected waves are strictly spherical. By moving the knife-edge at right angles to the optical axis, one can obtain a sudden, uniform extinction of the image at the instant it covers the light source image.

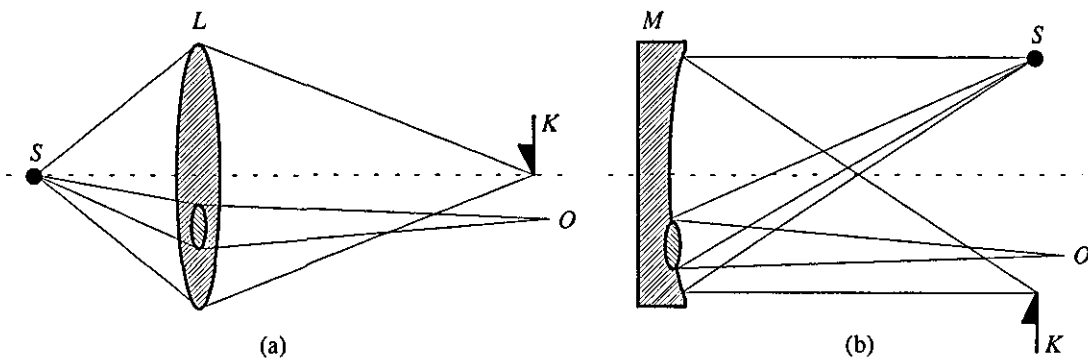


Fig. 1.5 - The principle of the schlieren method.

If some sections of the objective or mirror distort the wave-front, the light from these sections will produce an image of the luminous point at a different place. As a result, at the moment that the main image of the light source is covered by the knife-edge, and the entire field is dark, some light from the optical non-homogeneity will nevertheless reach the observer's eye placed at O . The non-homogeneity will appear illuminated against a dark background. When the knife-edge is rotated through 180° and moved toward the optical axis, the non-homogeneity will appear dark against a bright background.

In real optical systems, different field points deflect the light through different angles. If the disturbance being studied is theoretically split into a large number of areas, each of these areas forms an elementary image source. Each area is extinguished at the moment that the knife-edge covers the respective light source image. For example, glass defects and surface finish inaccuracies become visible and the object appears as if it were illuminated from the side.

If the point light source is replaced by an extended source, the source image is covered by the knife-edge gradually and not instantaneously. The illumination caused by field points deflecting the light through different angles will then vary and a gradual change

in the illumination will be observed corresponding to the change in the angle of light deviation in the disturbance. The sensitivity of this method depends on the light source. By using a sufficiently small light source, one can obtain a much greater difference between field point illuminations.

Since the source image shift depends on the angle of deviation in the non-homogeneity, the change in the light wave shape caused by the object can be estimated from the distribution of illumination in the schlieren pattern. If the knife-edge is used as the viewing diaphragm, the image illumination is affected only by shifts of the source image perpendicular to the knife-edge. If the image shifts along the knife-edge, there is no change in illumination. A schlieren instrument with a knife-edge therefore makes it possible to distinguish between points that differ in the value of the projections of the angle of light deviation on the plane perpendicular to the knife-edge.

The replacement of the point light source by a slit source [14] marked a significant advance, making it possible to obtain a much larger light flux without loss of sensitivity. By increasing the light flux, one reduces the exposure time and can study in detail various high-speed processes, such as the flight of a projectile, the motion of a shock wave, etc. The reduction of the exposure time is also helpful in steady-state processes, since the errors caused by variations of the instrument calibration with time due to fluctuations in temperature, atmospheric pressure, etc., are correspondingly reduced.

Qualitative schlieren photography yields valuable information on optical disturbances. Schlieren photography can, however, also be used for determining the shape of the light wave-front after its passage through the non-homogeneity, and therefore for establishing the refractive index distribution inside the non-homogeneity. Besides the schemes shown in Fig. 1.5, which are applicable for qualitative studies, other schemes with the non-uniformity being investigated contained in a collimated beam of light can be used too. These schemes best provide for quantitative measurements because it is possible to define the dimensions of the area that is studied in a direction perpendicular to the passage of the light rays, as shown in Fig. 1.6.

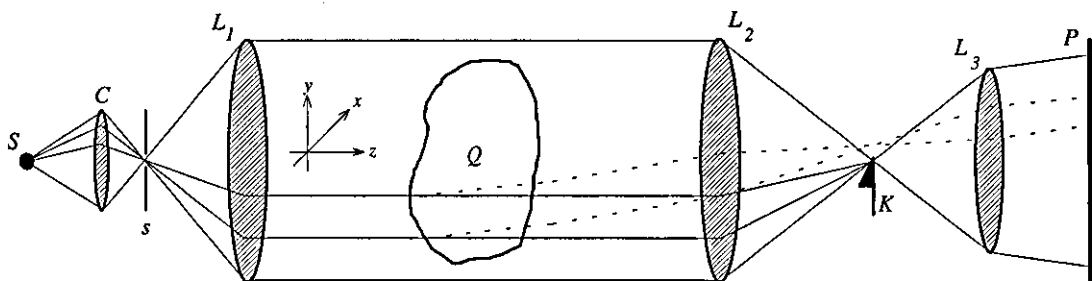


Fig. 1.6 - The arrangement of a typical schlieren apparatus.

In this case, the light from the source S is collected by the condenser C in the plane of the slit s which is located in the focal plane of a long-focus collimating lens L_1 . The non-uniformity Q to be studied is contained in the collimated beam of light that is beyond the first lens. The objective lens L_2 , situated in the collimated beam, forms an image of the slit in the plane of the knife-edge K and, together with an auxiliary lens L_3 , forms the image of the disturbance on the screen P .

The schlieren methods depend on the deflection of a ray of light from its undisturbed path when it passes through a medium in which there is a component of the gradient of refractive index normal to the ray. It may be shown [12,13,15] that the curvature of the ray is proportional to the refractive index gradient in the direction normal to the ray. According to Fig. 1.6 the curvatures in the xz and yz planes respectively are given by

$$\frac{\partial^2 x}{\partial z^2} = \frac{1}{n} \frac{\partial n}{\partial x} \quad (1.3-1)$$

and

$$\frac{\partial^2 y}{\partial z^2} = \frac{1}{n} \frac{\partial n}{\partial y}, \quad (1.3-2)$$

where n is the refractive index. If the total angular deflection in the xz and yz planes are denoted by ϵ_x and ϵ_y respectively, then

$$\epsilon_x = \int \frac{1}{n} \frac{\partial n}{\partial x} dz \quad (1.3-3)$$

and

$$\epsilon_y = \int \frac{1}{n} \frac{\partial n}{\partial y} dz, \quad (1.3-4)$$

where the integrals are taken over the width of the disturbance and the deflection being in the direction of the refractive index gradient, that is, towards the region of highest density. In the schlieren method, the deflection of the rays and the illumination are proportional to the first derivative of the change of the refractive index in the perturbed medium, that is, in the general case for a circular source, the change of illumination of the schlieren image is subject to the law:

$$\Delta I = K \int \left(\frac{\partial n}{\partial x} + \frac{\partial n}{\partial y} \right) dz, \quad (1.3-5)$$

where K is a constant, x and y are co-ordinates in a plane normal to the ray of light, and z is the co-ordinate in the direction of the ray. Integration makes it possible to convert directly to the refractive index of the area of the non-uniformity being studied.

Figure 1.7 shows, as an illustrating example, a schlieren image of a focused shock wave travelling in water, where the bright areas correspond to different refractive index gradients generated in the water due to changes in the pressure through the medium.



Fig. 1.7 - Schlieren image of a focused shock wave in water.

Schlieren images can change depending on the shape of the visualization diaphragm or stop. The variants applied are a knife-edge (a plate with a sharp straight edge), a point (a circular opaque disc), a circular aperture, thread, slit, line grating, point grating, or curvilinear diaphragm. Two positions of the diaphragm or stop are also possible: on the focal point or out of focal point on the objective lens. In addition, schlieren methods are distinguished by the arrangement of the visualizing diaphragm in the optical system and also by the shape and position of the light source. The above conditions may occur in various combinations, which means that in practice there is a wide variety of different schlieren arrangements that can be implemented.

The measurable quantity for a schlieren instrument is the angle φ of the light deflection in a non-homogeneous medium. In a system with a collimated beam of light (Fig. 1.6), the deflection of a ray of light in the focal plane of the objective lens L_2 will be given by $a = f \tan \varphi$, where f is the focal length of the objective lens, or, since the angle φ is small, by a value $a = f\varphi$. The sensitivity of the instrument is defined as the ratio of the value of the change in the reading instrument and the change of the parameter being measured. For a schlieren instrument, these quantities depend on the arrangement implemented and applied.

The effects of a displacement of the image of the source on the illumination at the viewing screen have been discussed based on geometrical optics. A more rigorous

examination based on the wave theory shows [16], however, that the effects of diffraction may be important. It is found that, although the illumination on the screen is usually approximately constant in the absence of the knife-edge, this is no longer the case when a knife-edge is present. The effects of diffraction then usually result in increased illumination on the screen as the boundaries of the aperture are approached. Also, the illumination does not fall sharply to zero at the boundaries, but dies away gradually. The amplitude and extent of these effects increase as the proportion of the image of the source cut-off by the knife-edge is increased. For the amount of cut-off used in most practical systems the effects are normally confined to regions very close to the images of the boundaries and result in the formation of narrow bright bands running adjacent to them.

If a deflection is produced in the parallel light beam by the introduction of a density gradient, the effects of diffraction may result in a change of illumination on the viewing screen which differs appreciably in both amplitude and extent from that expected on the basis of geometrical optics. These effects are usually small if the apparatus is used at low sensitivity with a small amount of the source image being cut-off, but may become dominant if the sensitivity is high. The effects are not readily calculable, and thus affect the accuracy of the schlieren technique when used for quantitative measurements. For this reason, the schlieren technique is normally used to measure big and sharp changes in the density of the medium, which correspond to large deflections of the incident illumination, with relatively low sensitivities. It is very difficult to measure small and smooth changes in the density of the medium due to diffraction effects and also due to practical limitations on the instrument such as positional accuracy of the diaphragm or stop.

1.4 Interferometric Techniques

The interferometric techniques, as a complement to the schlieren technique, are applied in the study of small and smooth changes in the density in inhomogeneous media, making it possible to obtain quantitative values of the refractive index of the media being studied. Subsequent calculation will give the values of the density of the substance, pressure, temperature of the medium, etc.

The basic design of stellar and spectral interferometers was given in 1881 by Michelson [17,18] in connection with experiments to measure the star diameters and the wavelengths of spectral lines. Today, these two methods are seen as the prototype of

important techniques in modern interferometry. The principles involved are as important today as they were then. To introduce the use of interference effects as an indirect way to perform measurements, the phenomenon of interference must be described.

Light is composed of electric and magnetic fields oscillating at frequencies of the order of 10^{15} Hz and propagating in space at a speed $c = c_0/n$, where $c_0 = 2.998 \times 10^8$ m/s is the speed of the light in vacuum and n is the refractive index of the medium. For a monochromatic and linearly polarized plane wave of light the optical field is fully described by the magnitude and phase distributions. Therefore, for a generic point p , a complex expression $U(p)$ can be used to represent the electric or magnetic component of light as

$$U(p) = |U(p)| \exp[-j\phi(p)], \quad (1.4-1)$$

where $|U(p)|$ and $\phi(p)$ are the magnitude and phase of the light wave respectively. Since the optical field oscillates at an extremely high rate, light detectors can measure only the time-averaged value of the optical power, namely irradiance or intensity. If ϵ is the permittivity of the medium, and indicating with an asterisk the complex conjugate, the intensity $I(p)$ can be expressed by

$$I(p) = \epsilon c U(p) U^*(p). \quad (1.4-2)$$

Interference occurs at a point p when two monochromatic light waves of the same frequency and polarization are superimposed. In this case, the resulting intensity produced by the sum of the two optical fields is not the sum of the individual intensities. If U_1 and U_2 are the considered fields at a point p , the total intensity I is proportional to the squared sum of the complex amplitudes and is given by

$$I = \epsilon c (U_1 + U_2)^2 = \epsilon c (U_1 + U_2)(U_1 + U_2)^*, \quad (1.4-3)$$

and finally

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2} \cos(\phi_1 - \phi_2). \quad (1.4-4)$$

It is worth noting that the third term in Eq. (1.4-4) contains the difference of the optical phases. Thus the difference in time or equivalently the optical phase difference can be evaluated indirectly from the intensity profile obtained by interference. The interference phenomenon is very clear when the two optical waves have the same intensity. In this

case, if the phase difference is an integer multiple of 2π there is constructive interference and the total intensity is four times that of a single wave. On other hand, destructive interference occurs when the phase difference is an integer multiple of π , and the corresponding intensity is zero.

Optical interference can be observed only if certain conditions are satisfied. The strongest constraint is the temporal and spatial coherence of the light [19]. Conventional light sources emit a sequence of photon wave trains whose electromagnetic fields oscillate with a continuous phase function only for a small time interval corresponding to the duration of an atomic transition. Therefore the optical wave is interrupted so frequently that any interference pattern cannot effectively be detected. This effect is closely related to the degree of monochromaticity of the light wave.

If it were possible to freeze in space the wave profile, on average the optical field would appear as a continuous sinusoidal function over an interval, which is called the coherence length. An effective way to produce interference at a point p is to split the light wave from the source into two beams, whose paths to p differ less than the coherence length. By doing so the waves superimposing at p belong to the same wave train emitted by the source and a stable interference pattern results.

Another property of the light that plays a fundamental role in the phenomenon of interference is polarization [20]. If two light waves are linearly polarized in two different states, the interference term of the intensity resulting from their superposition contains the cosine of the angle between the two polarization planes. In fact when squaring the sum of two vectors a scalar product is obtained corresponding to the interference term. Therefore, the maximum interference occurs if the two waves have the same polarization whilst no interference occurs if the polarization vectors are perpendicular.

Nowadays, these interferometric techniques have been highly developed. As a result, a large number of different instruments and interferometric apparatus have been implemented for a variety of applications which can be found in literature [21].

The optical scheme of the most widely used interferometer is shown in Fig. 1.8. It is a Mach-Zehnder type interferometer where the light from the source S is collected by the condenser C in the plane of the slit s , which is located in the focal plane of the collimating lens L_1 , to produce a collimated beam of light that passes through the beam-splitter BS_1 which splits it into two collimated beams. One of these beams continues in

the direction of the mirror M_1 and the other in the direction of the mirror M_2 . Moreover, the second beam passes on its way through the non-uniformity being studied Q . Further, both the beams are recombined by means of the second beam-splitter BS_2 . Since the second beam changed its velocity in passing through the non-uniformity [3], it acquires a phase shift. In other words, this phase shift is due to the optical path difference in the two arms of the interferometer introduced by the disturbance under study. Then the two beams will interfere and form an interference fringe pattern which can be viewed on the screen or photographic film P using the auxiliary objective lens L_2 which images the region of interest Q .

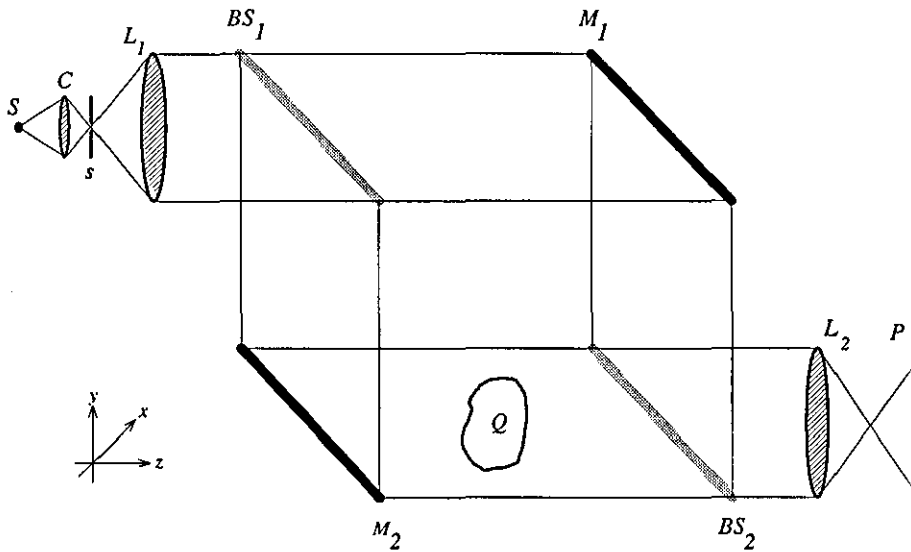


Fig. 1.8 - Schematic diagram of the Mach-Zehnder interferometer.

The refractive index at any point in the medium can be defined by the shape of the observed interference fringes and the distance between them.

The combined use of schlieren and interference instruments holds much promise. Since interferometers are sensitive to changes in the optical length of the light ray's path, they possess a great advantage over the shadow and the schlieren instruments in allowing quantitative measurements of the refractive index of a medium. High sensitivities are normally achieved, and highly accurate measurements can be performed with these interferometers. On the other hand, schlieren methods are sensitive to the change of the gradient of the optical length of the path. Therefore, by means of schlieren instruments, it is often more convenient to determine the shape and location of such gradients instead of performing accurate measurements.

Figure 1.9 shows, as an illustrating example, an interferometric image of a laser breakdown in air produced in one of the arms of a Mach-Zehnder interferometer. The shift of

the interference fringes due to the refractive index changes in the medium are clearly visible and phase measurements can also be performed from such an interferogram.

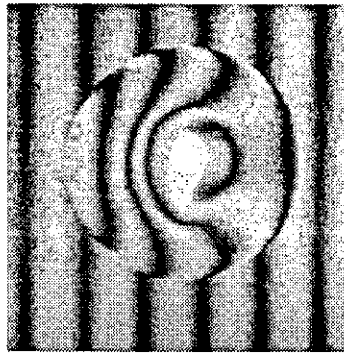


Fig 1.9 - Interferometric image pattern created by a laser break-down.

These interferometric techniques and the corresponding interferometric instruments always use as the source of illumination monochromatic light with wavelength λ . Lasers are widely used as sources of monochromatic light in interferometers [21-23], which makes it possible to produce perfect interferograms due to the good coherency of the laser light. A further advantage is the ability to generate extremely short pulses of light.

1.5 An Alternative to Interferometric Techniques

Since shadowgraphy, schlieren, and interferometric techniques give observations which depend on the refractive index (density) and on different derivatives of the refractive index, these three methods are complementary rather than alternative, each showing features which may not be so clearly observed by the other.

Of the three methods, the only one which has been used extensively for quantitative measurements is interferometry, and the main use of shadowgraphy and the schlieren methods is at present to show the positions and shapes of regions of refractive index changes. Used in this fashion these last two methods require only comparatively simple and inexpensive apparatus, but frequently they give information that is of great value in understanding the event being investigated. Interferometric methods require complex and often sophisticated apparatus and their practical implementation is not always simple. Another aspect to be considered is the cost of the optical components for implementing a very good and accurate instrument, which is normally high.

In this section a more simple and inexpensive method is proposed as an alternative to the complex and expensive interferometric techniques. The method is based on Fourier

transforms and involves the recovery of phase information from intensity information associated with an optical wave disturbance using special algorithms. As discussed before, knowing the phase it will be possible to assess the refractive index changes, density, etc., of the medium under study.

1.5.1 Historical Review

It is in the appreciation of the phenomenon of wave interference that the concept of phase becomes important quantitatively. Consequently, Thomas Young must be considered the *father of phase*, despite the celebrated controversy of a hundred years earlier over Newton's and Huygen's rival views of the nature of the light [24].

The dichotomy between frequency and space domains had negligible practical impact before this century. Up until then all image had been formed directly in the focal planes of optical instruments. This changed with Michelson's interferometric investigations [25]. Even though it was soon realized that interpretations of the fringe depth in interferograms could be ambiguous, and no clear statement of a phase problem came until X-ray crystallography had fully absorbed and outgrown the initial insights of the founders of the science. It seems that crystallographers only came to realize the full technical significance of the consequences of it being effectively impossible to measure the phase of a X-ray beam after the Fourier theoretical basis of their science had been widely promulgated in the decade before World War II.

Measurement science involves two very different classes of the phase problem. The first relates to periodic images, such as occur in crystallography, especially of the X-ray variety, but also in electron microscopy and neutron diffraction. The second class of phase problems concerns images of finite size, such as are of interest in astronomy, radio-engineering, often in microscopy, and throughout most of image science in general.

For the first class of problems, an enormous effort has been spent on estimating crystallographic phases to make possible the determination of crystal structures. Crystallographic phase problems of ever increasing difficulty and complication are continuously yielding to combinations of ingenious mathematics, complex experimental techniques, and tenacious computation. Most of the techniques for phase recovery in this area have been developed by Beurger, Ramachandran, and Woolfson [26-28]. Also,

the work carried out by the electron microscopists Hawkes and Misell [29,30] has an important role on this particular problem of phase retrieval.

Something very important that crystallographers seem to have been the first to notice is that phase tends to dominate intensity, as Ramachandran and Srinivasan [27] illustrated in a crystallographic context. Oppenheim and Lim [31] have emphasised this recently in a standard image processing context. As a consequence there has been noticeable initial resistance within the image processing community to recent claims that a very wide class of phase problems do actually possess effectively unique solutions. Nevertheless, these claims now seem to be supported with evidence.

Of all the established imaging technologies, the one with the best theoretical resolution is electron microscopy. The theoretical resolution limit is never approached closely in practice because of technical difficulties associated with collimating the electromagnetic lens system. From certain measurements, possible under particular circumstances, one can estimate the phase distortion present in the back-focal plane, implying that phase correction should give improved images. Since it is impossible to sense electron phase, there is a phase problem. Remarkably enough, this served as the inspiration for holography, which Gabor [32] invented with the idea of enhancing electron micrographs by subsequent optical processing. Even though electron holography has been worked on steadily for more than 40 years, it has had virtually no impact on the practice of microscopy, whereas optical holography is now an important technical science [33].

While optical approaches to the phase problems of electron microscopy have not initiated significant technical advances, digital approaches have had their success [34], most notably perhaps by invoking what is called the Gerchberg and Saxton algorithm [35]. In many practical applications, measurements have to be made in the Fourier domain, also called the frequency or spectral domain. So, the true image has to be inferred from its Fourier spectrum and for that reason several digital techniques based on Fourier transforms have been proposed and developed by Bates and Fienup [36-38].

The phase problems that most often arise in communications [39] and radio-engineering [40] are one-dimensional. Optical scientists have also paid much attention to the one-dimensional phase problems, presumably to avoid what may have seemed to them at the time the unnecessary complications of multi-dimensional analysis. It is well known that one-dimensional phase problems only possess unique solutions in very special cases [41]. For about two-dozen years it has been noticed, from computational experimentation with two-dimensional images, that the image form usually appears to

be uniquely related to its Fourier intensity. Although an early example [42] involved fairly conventional Fourier processing, the first comprehensive results were obtained by the maximum entropists [37]. Because of the importance of image positivity to the uniqueness question, one can perhaps see more clearly now why maximum entropy algorithms can be so powerful: because of the presence of the logarithm in the various functions that have been invoked. Since the logarithm changes abruptly from real to complex as its real argument passes through zero, the maximum entropy algorithms strongly inhibit the appearance of negative parts in reconstructed images. There was a tendency at first to try and explain these results on the basis of the images being somehow special [43], particularly since maximum entropy is known to sometimes come up with images forms closer to the auto-correlation of the image than the image itself. However, as the number of special cases multiplied, due even more to Fienup's efforts [38,44] than to those of the entropists, this explanation became less and less credible.

Once it is widely appreciated that the form of two-dimensional, positive (that is, real and non-negative) images can be recovered unambiguously from the intensities of their Fourier transforms, a steady stream of important applications can be expected. The remarkable difference between the one-dimensional and the two- or multi-dimensional phase problems was clearly established by Taylor, Bates, Bruck, Fienup, and Hayes [41,45-48] which states that solutions to the one-dimensional case are almost never unique whereas solutions to the latter are almost always unique. Bruck and Sodin [46] pointed out that multi-dimensional polynomials are usually unfactorizable, thereby implying that multi-dimensional phase problems can be expected to have unique solutions except for images possessing special symmetries. This gave solid support to Fienup [44] and to other scientists [49,50], and it helped to inspire the Fienup algorithm a version of which is described in [51].

Algorithms that have proved their worth in the recovery of two-dimensional phase distributions are adaptations [38,44,47] of the Gerchberg and Saxton algorithm [35]. Unadulterated, these algorithms can be ineffective, especially when used with measured data contaminated with appreciable amounts of noise [52]. Appropriately orchestrated algorithms and variations show promise of successful phase recovery in widening ranges of situations of practical interest.

There are many classes of problems that can either be transformed so as to be cast as pure phase problems or are to be thought of as restricted phase problems in that *a priori* information is available. The emphasis in this thesis is on the pure phase problems in

forms of two-dimensional, complex valued images associated with optical disturbances. In order to recover the phase for complete reconstruction of the optical wave, and bearing in mind the historical review of the phase problem and the techniques available to overcome such problems, a technique is proposed in the next section.

1.5.2 The Proposed Technique

The main problem in astronomy and in most of image science is to reconstruct and generate images from measured spectral intensities [38,48,52]. To form useful images about an object disturbance then, it will be necessary to develop versatile computational procedures to retrieve information for inferring uniquely the spectral phase from sufficiently closely spaced samples of measured spectral intensities in an optical system [42,46,51]. Several constraints are in general applied to the phase retrieval process, and in particular it is assumed that the object image is real and non-negative [45,53-55].

The problem considered here is quite different because it concerns the reconstruction of optical wave-fronts and therefore the reconstruction of object images that are complex valued. Since these optical wave-fronts and object images are complex functions they are characterized by a magnitude and phase in the spatial domain. Unlike the astronomy problem, and in this case in particular, spatial and spectral intensities can be made available on an optical system and accurate intensity measurements can also be made.

Recalling the arrangement of a typical schlieren apparatus, shown in Fig. 1.6 and repeated here in Fig. 1.10 for convenience, the proposed technique concerns the measurement of the intensity distributions in the Fourier or spectral domain, FD, and in the image or spatial domain ID when an object disturbance is placed in the object domain OD of the optical system shown.

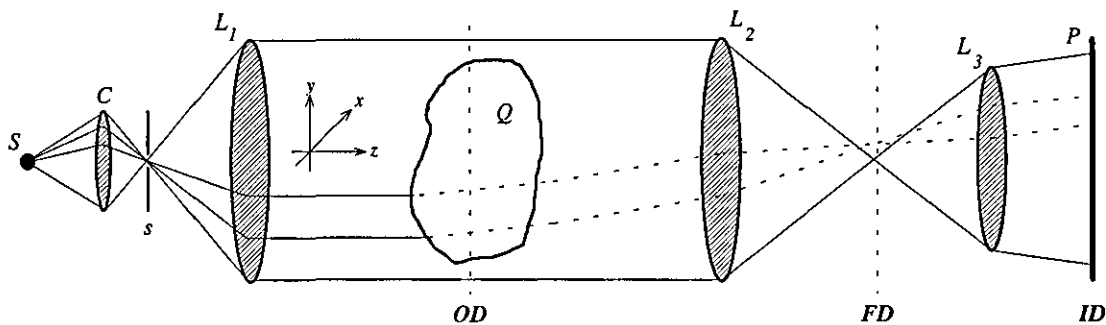


Fig. 1.10 - Optical apparatus for spectral and spatial intensity measurements.

Then, by proper digitalization the measured information can be stored and processed in a digital computer. The phase problem resumes to retrieve the phase information that uniquely fills both the spatial and spectral constraints. In particular an algorithm based on the Gerchberg and Saxton algorithm [35] and its modifications [56] will be intensively used to retrieve phase information from the measurements of spatial and spectral intensities. The algorithm is presented and discussed in chapter four. Computer simulations and experimental results obtained with this technique are presented and discussed in chapter five. The generation of schlieren images on the computer screen using the retrieved phase information is also discussed in chapter six for both simulated and real cases.

1.6 Summary

The principal purpose of this chapter has been to present a general view of the available methods and techniques for qualitative and quantitative characterization of optical disturbances. Linear displacement, angular deflection, and phase difference of light rays are the three quantities that form the basis of three methods, namely, shadowgraphy, schlieren and interference, that have been briefly described and compared in the context of optical measurements of refractive index (density) in a given medium.

A review of phase retrieval problems in different areas of science and techniques to overcome such problems has also been presented. An alternative to the interferometric techniques has been proposed as a means of phase recovery from single intensity measurements in an optical system. The technique, which is based on the Gerchberg and Saxton algorithm, will be used for wave-front reconstruction of optical disturbances.

References

1. H. A. Lorentz, "Over het Verband Tusschen de Voortplantingssnelheid van het licht en de Dichtheid en Samenstelling der Middenstoffen", *Ann. Phys. Chem.* **9**, 641-665 (1880).
2. J. H. Gladstone and T. P. Dale, "Researches on the Refraction, Dispersion and Sensitiveness of Liquids", *J. Phil. Trans.* **153**, 317-343 (1863).
3. F. A. Jenkins and H. E. White, *Fundamentals of Optics*, 4th Edition, McGraw-Hill Ed., Singapore (1981).
4. A. S. Dubovik, *The Photographic Recording of High-Speed Processes*, John Wiley & Sons, Inc., New York (1981).
5. V. Dvorak, "Über eine neue Einfache art der Schlierenbeobachtung", *Ann. Phys. Chem.* **9**, 502 (1880).
6. P. H. Cords, W. R. Noyes, and P. A. Thurston, "A Modern Shadowgraph System for use in High Velocity Aeroballistic Ranges", *Proc. of the 5th International Congress on High-Speed Photography*, Washington (1960).
7. N. F. Barnes and S. F. Bellinger, "Schlieren and Shadowgraph Equipment for Air Flow Analysis", *J. Opt. Soc. Am.* **35**, 497-509 (1945).
8. L. M. Foucault, "Description des Procèdes Employes pour Reconnaître la Configuration des Surfaces Optiques", *C. R. Acad. Sci. Paris* **47**, 958 (1858).
9. L. M. Foucault, "Memoire sur la Construction des Telescopes en Verre Argente", *Ann. Obs. Imp. Paris* **5**, 197 (1859).
10. A. Töpler, *Beobachtungen nach einer neuen Optischen Methode*, Cohen & Sons, Bonn (1864).
11. L. Rayleigh, "On Methods for Detecting Small Optical Retardations, and on the Theory of Foucault's Test", *J. Phil. Mag.* **33**, 161 (1917).

12. H. Shardin, "Die Schlierenverfahren und ihre Anwendungen", *Ergebn. Exakt. Naturw.* **20**, 303 (1942).
13. L. A. Vasil'ev, *Schlieren Methods*, Keter Inc., New York (1971).
14. D. D. Maksutov, "Schlieren Methods in the Study of Optical Systems", *Problems in Advanced Physics XXIII*, State Publishing House for Technical and Theoretical Literature, Russia (1934).
15. D. W. Holder and R. J. North, *Schlieren Methods*, Her Majesty's Stationery Office, London (1963).
16. H. J. Schafer, "Physical Optic Analysis of Image Quality in Schlieren Photography", *J. Soc. Motion Picture Engineers* **53**, 524 (1949).
17. A. A. Michelson, "On the Application of Interference Methods to Astronomical Measurements", *J. Phil. Mag.* **30**, 1-21 (1890).
18. A. A. Michelson, "On the Application of Interference Methods to Spectroscopic Measurements", *J. Phil. Mag.* **31**, 338-346 (1891).
19. A. S. Marathay, *Elements of Optical Coherence Theory*, John Wiley & Sons, Inc., New York (1982).
20. E. Hecht, *Optics*, 2nd Edition, Addison-Wesley Publishing Company, Inc., (1987).
21. D. Malacara, *Optical Shop Testing*, 2nd Edition, John Wiley & Sons, Inc., New York (1992).
22. M. Hugenschmidt and K. Vollrath, "Interferometry of Rapidly Varying Phase Objects Using the Fundamental and the Harmonic Wavelengths of a Ruby Laser", *Proc. of the 9th International Congress on High-Speed Photography*, Denver (1970).
23. V. Sernas and J. A. Jones, "A Wollaston Prism Schlieren Interferometer", *Proc. of the 9th International Congress on High-Speed Photography*, Denver (1970).

24. I. Asimov, *Biographical Encyclopedia of Science and Technology*, Doubleday, New York (1964).
25. D. H. De Vorkin, "Michelson and the Problem of Stellar Diameters", *J. History of Astronomy* **6**, 1-18 (1975).
26. M. J. Beurger, *Vector Space*, John Wiley & Sons, Inc., New York (1959).
27. G. Ramachandran and R. Srinivasan, *Fourier Methods in Crystallography*, John Wiley & Sons, Inc., New York (1970).
28. M. M. Woolfson, *Direct Methods in Crystallography*, Oxford University Press, Oxford (1961).
29. P. W. Hawkes, *Computer Processing of Electron Microscope Images*, Springer-Verlag, Berlin (1980).
30. D. L. Misell, *Image Analysis, Enhancement and Interpretation*, North-Holland, Amsterdam (1978).
31. A. V. Oppenheim and J. S. Lim, "The Importance of Phase in Signals", *Proc. of the IEEE* **69**, 529-541 (1981).
32. J. D. Gabor, "A New Microscopic Principle", *Nature* **161**, 777-778 (1948).
33. H. J. Caulfield, *Handbook of Optical Holography*, Academic Press, New York (1979).
34. D. L. Misell, "The Phase Problem in Electron Microscopy", in *Advances in Optical and Electron Microscopy* (V. E. Cosslett and R. Barer Ed.) **7**, 185-279 (1978).
35. R. W. Gerchberg and W. O. Saxton, "A Practical Algorithm for the Determination of Phase from Image and Diffraction Plane Pictures", *Optik* **35**, 237-246 (1972).
36. R. H. T. Bates, "On Phase Problems II", *Optik* **51**, 223 (1978).

37. R. H. T. Bates, "Astronomical Speckle Imaging", *Physics Reports* **90**, 203-297 (1982).
38. J. R. Fienup, "Reconstruction of an Object from the Modulus of its Fourier Transform", *Optics Letters* **3**, 27-29 (1978).
39. B. P. Lathi, *Communication Systems*, John Wiley & Sons, Inc., New York (1968).
40. A. C. Schell, "The Selection and Evaluation of Antennas for Surveying Incoherent Source Distributions" in *Electromagnetic Theory and Antennas* (E. C. Jordan Ed., Pergamon Press, Oxford), 1233-1248 (1963).
41. L. S. Taylor, "The Phase Retrieval Problem", *IEEE Trans. on Antennas & Propagation* **AP-29**, 386-391 (1981).
42. P. J. Napier and R. H. T. Bates, "Inferring Phase Information from Modulus Information in Two-dimensional aperture Synthesis", *Supplement Series of Astronomy and Astrophysics* **15**, 427-430 (1974).
43. R. H. T. Bates, "Fringe Visibility Intensities May Uniquely Define Brightness Distributions", *Astronomy and Astrophysics* **70**, 27-29 (1978).
44. J. R. Fienup, "Phase Retrieval Algorithms: a Comparison", *Applied Optics* **21**, 2758-2769 (1982).
45. R. H. T. Bates, "Uniqueness of Solutions to Two-dimensional Fourier Phase Problems for Localized and Positive Images", *Computer Graphics, Vision & Image Processing* **25**, 205-217 (1984).
46. Y. M. Bruck and L. G. Sodin, "On the Ambiguity of the Image Reconstruction Problem", *Optics Communications* **30**, 304-308 (1979).
47. J. R. Fienup, "Space Object Imaging through the Turbulent Atmosphere", *Optical Engineering* **18**, 529 (1979).
48. M. H. Hayes, "The Reconstruction of a Multi-dimensional Sequence from the Phase or Magnitude of its Fourier Transform", *IEEE Trans. on Acoustics, Speech & Signal Processing* **ASSP-30**, 140-154 (1982).

49. A. H. Greenaway, "Proposal for Phase Recovery from a Single Intensity Distribution", *Optics Letters* **1**, 10-12 (1977).
50. A. H. Greenaway, "The Phase Problem in Astronomy", *J. Optics* **10**, 308-310 (1979).
51. R. H. T. Bates, "Fourier Phase Problems are Uniquely Solvable in more than One Dimension I: Underlying Theory", *Optik* **61**, 247 (1982).
52. R. H. T. Bates and W. R. Fright, "Reconstructing Images from their Fourier Intensities", in *Advances in Computer Vision & Image Processing* (T. S. Huang Ed., Jai Press, Inc., London) **1**, 227-264 (1984).
53. M. A. Fiddy, B. J. Brames, and J. C. Dainty, "Enforcing Irreducibility for Phase Retrieval in Two Dimensions", *Optics Letters* **8**, 96 (1983).
54. R. H. T. Bates and M. J. McDonnell, *Images Restoration and Reconstruction*, Oxford University Press, Oxford (1986).
55. J. C. Dainty and J. R. Fienup, "Phase Retrieval and Image Reconstruction for Astronomy", in *Image Recovery: Theory and Application* (H. Stark Ed., Academic Press, Inc., San Diego) **7**, 231-275 (1987).
56. G. Yang, B. Dong, B. Gu, *et. al.*, "Gerchberg-Saxton and Yang-Gu Algorithms for Phase Retrieval in a Nonunitary Transform System: a Comparison", *Applied Optics* **33**, 209-218 (1994).

CHAPTER 2

A Digital Image Processing System

2.1 Introduction

What is digital image processing? Image processing is a science that deals with images and image data. It covers a broad spectrum of techniques that are applicable to a wide range of applications. Image processing can be thought of as a special form of two-dimensional signal processing used to uncover information about images. In general, image processing techniques are applied to images or image data when:

- enhancement or modification of an image is necessary to improve the appearance or to highlight some aspect of the information contained in the image;
- elements within an image need to be characterized, classified, matched or measured;
- portions of images need to be combined or image elements need to be reorganized;
- information contained within an image need to be recovered or reconstructed.

The techniques for image processing can be applied to data even if the data is not in the visible form. The manipulation of visible data is just one of the many uses of image processing - probably the predominant one. Image processing can be used to produce a visible image of purely numeric data enhanced in some manner to highlight some aspect of the data. Applying an image processing transformation to an image is not always done with the appearance of the image in mind. Actually, the result might not be pleasing to look at. Aesthetics are not the only criterion to judge the effectiveness of the applied transformation. If the transformation is designed to bring out additional information and/or details not visible in the original image, the result can be considered successful even if it is not pleasing to look at. These processing results are intended for human interpretation. However, if the processing results are intended for autonomous machine perception a different area of application is defined. In this case, interest is focused on procedures for extracting from an image information in a form suitable for computer processing. This area is then referenced by digital image processing instead of image processing. So, digital image processing concerns the acquisition, discretization, storage, and interpretation of information in the form of a digital image by the computer.

Many different image processing transformations and algorithms can be developed and implemented which can be applied to an image individually or in conjunction with other

functions. However, the application of these transformations and algorithms is not commutative in nature. That is, the order of application is very important to achieve the desired effect. Normally, they can be utilized in a cook-book manner without having to understand completely their mathematical underpinnings. Of course, the results obtained will be more controllable if the mathematics behind the transformations is fully understood.

The number of different applications and groups interested in such applications is growing. It is probably safe to say that there is virtually no field where images are acquired, which means virtually every branch of science, that does not have active or potential work going on or being planned in digital image processing. As examples, the following can be considered, medicine (diagnostic X-rays, cell biology, anatomy, physiology); physics (plasma diagnostics, ultrahigh-pressure shock waves, solid-state phenomena, optical diagnostics, high-speed photography); non-destructive testing (visual quality control inspection, radiographic inspection, acoustic holography); weather forecasting (observation of visible cloud features from weather satellites); resource exploration (two-dimensional time/reverberation seismic signal processing); and so on.

The work presented in this chapter is mainly concerned with the implementation and development of a system for digital image processing in the Optoelectronics Laboratory in the Department of Physics at the Loughborough University of Technology. The idea is to use the developed system to support the research to be carried out, integrating all the fundamental and global tools for image manipulation and processing. Some specific techniques, transformations and algorithms specially implemented for particular applications will be included and future upgrades of the system will be supported.

Section 2.2 presents the digital image fundamentals. It includes an image model and an intuitive definition of image sampling and quantification. The principal elements in a basic digital image processing system are described in Section 2.3, and hardware and software considerations of the developed system are presented in Section 2.4. Finally, a brief summary is included in Section 2.5.

2.2 Digital Image Fundamentals

The term monochrome image or simply image, refers to a two-dimensional light intensity function $f(x,y)$, where x and y denote spatial co-ordinates and the value of f

at any point (x,y) is proportional to the brightness or grey level of the image at that point. It is sometimes useful to view an image function in perspective with the third axis being brightness.

A digital image is an image $f(x,y)$ that has been discretized both in spatial co-ordinates and in brightness, and it can be considered as a matrix whose row and column indices identify a point in the image and the corresponding matrix element value identifies the grey level at that point. The elements of such a digital array are called image elements, picture elements, pixels or pels, with the last two names being commonly used abbreviations of picture elements.

2.2.1 The Image Model

As previously defined, the term image refers to a two-dimensional light intensity function denoted by $f(x,y)$. Since light is a form of energy, $f(x,y)$ must be non-zero and finite, that is,

$$0 < f(x,y) < +\infty. \quad (2.2-1)$$

The images perceived in everyday visual activities normally consist of light reflected from objects. The basic nature of $f(x,y)$ may be considered as being characterized by two components. One component is the amount of source light incident on the scene to be viewed, while the other is the amount of light reflected by the objects in the scene. These components are appropriately called the illumination and reflectance components [1], and are denoted by $i(x,y)$ and $r(x,y)$, respectively. The functions $i(x,y)$ and $r(x,y)$ combine as a product to form $f(x,y)$:

$$f(x,y) = i(x,y)r(x,y), \quad (2.2-2)$$

where

$$0 < i(x,y) < +\infty \quad (2.2-3)$$

and

$$0 < r(x,y) < 1. \quad (2.2-4)$$

Equation (2.2-4) indicates the fact that reflectance is bounded by 0 (total absorption) and 1 (total reflectance). The nature of $i(x,y)$ is determined by the light source, while the nature of $r(x,y)$ is determined by the characteristics of the objects in a scene.

The values given in Eqs. (2.2-3) and (2.2-4) are theoretical bounds. The intensity of the resulting monochrome image f at co-ordinates (x,y) will be called the grey level l of the image at that point. From Eqs. (2.2-2) through (2.2-4), it is evident that l lies in the range

$$L_{\min} \leq l \leq L_{\max}. \quad (2.2-5)$$

In theory, the only requirement on L_{\min} is that it be positive, and on L_{\max} that it be finite. In practice, $L_{\min} = i_{\min} r_{\min}$ and $L_{\max} = i_{\max} r_{\max}$. The interval $[L_{\min}, L_{\max}]$ is called the grey level scale and it is common practice to shift this interval numerically to the interval $[0, L]$, where $l=0$ is considered black and $l=L$ is considered white in the scale. All intermediate values are shades of grey varying continuously from black to white.

To complete this image model the image degradation process should be included [2]. The degradation process will be modelled as an operator or system H , which together with an additive noise term $\eta(x,y)$ operates on an input image $f(x,y)$ to produce the degraded image $g(x,y)$ with the input-output relationship given by the expression

$$g(x,y) = Hf(x,y) + \eta(x,y). \quad (2.2-6)$$

The operator H , in general, can be a linear or non-linear operator and also it can be space-invariant or variant.

2.2.2 Sampling and Quantification

In order to be in a form suitable for computer processing, an image function $f(x,y)$ must be digitized both spatially and in intensity. Digitization of the spatial co-ordinates (x,y) will be referred to as image sampling, while intensity digitization will be called grey level quantification.

Suppose that a continuous image $f(x,y)$ is approximated by equally spaced samples arranged in a form of an $N \times N$ array as shown in Eq. (2.2-7), where each element of the array is a discrete quantity:

$$f(x,y) = \begin{bmatrix} f(0,0) & f(0,1) & \dots & f(0,N-1) \\ f(1,0) & f(1,1) & \dots & f(1,N-1) \\ \vdots & \vdots & \dots & \vdots \\ f(N-1,0) & f(N-1,1) & \dots & f(N-1,N-1) \end{bmatrix} \quad (2.2-7)$$

The right side of this equation represents what is commonly called a digital image, while each element of the array is referred to as an image element, picture element, pixel, or pel, as indicated in the beginning of Section 2.2. The terms image and pixels will be used throughout this thesis to denote a digital image and its elements. Digitization of an image need not be limited to square arrays. However, the following discussions will often be simplified by adoption of this convention.

The above digitization process requires that a decision be made on a value for N as well as on the number of discrete grey levels allowed for each pixel. It is common practice in digital image processing to let these quantities be integer powers of two, that is,

$$N = 2^n \quad (2.2-8)$$

and

$$G = 2^m, \quad (2.2-9)$$

where G denotes the number of grey levels. It is assumed in this section that the discrete levels are equally spaced between 0 and L in the grey scale. Using Eqs. (2.2-8) and (2.2-9) the number, b , of bits required to store a digitized image is given by

$$b = N \times N \times m. \quad (2.2-10)$$

For example, Fig. 2.1 shows a 512×512 digital image with 256 grey levels ($m = 8$). This

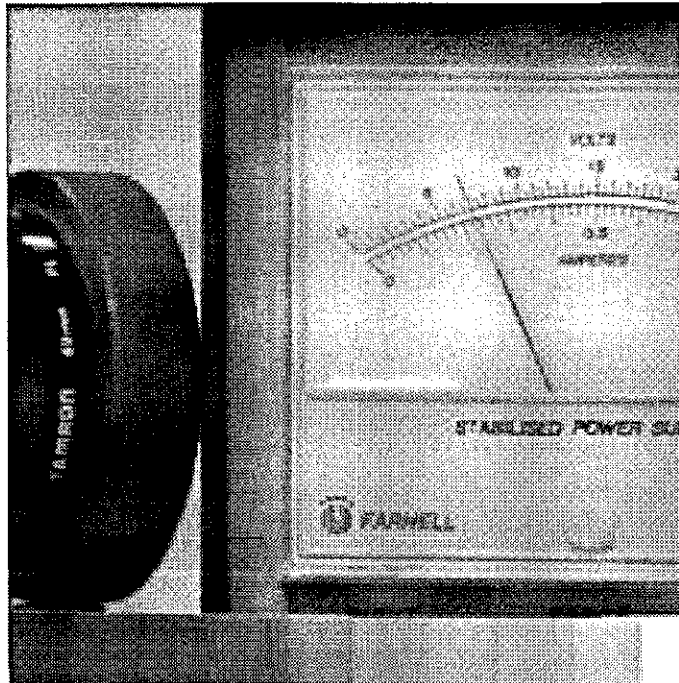


Fig. 2.1 - A 512×512 digital image with 256 grey levels.

image requires 2,097,152 bits of storage corresponding to 262,144 bytes or equivalently 256 kbytes.

Since Eq. (2.2-7) is an approximation to a continuous image, a reasonable question to ask at this point is how many samples and grey levels are required for a good approximation? The resolution, that is, the degree of discernible detail, of an image is strongly dependent on both N and m . The more these parameters are increased, the closer the digitized array will approximate the original image. However, Eq. (2.2-10) clearly points out the unfortunate fact that storage and, consequently, processing requirements increase rapidly as a function of N and m .

The number of samples and grey levels required to produce a faithful reproduction of an original image depends on the image itself. This problem will be the subject of a detailed analysis in the next chapter where the sampling theorem is discussed. As an intuitive basis for comparison [2,3], the requirements to obtain a quality comparable to that of monochrome TV pictures over a wide range of image types are on the order of 512×512 pixels with 128 grey levels ($m = 7$). In general and as a rule, a minimum system for general image processing work should be able to display 256×256 pixels with 64 grey levels ($m = 6$).

2.3 Elements of a Digital Image Processing System

The components of a basic, general purpose digital image processing system are shown in Fig. 2.2. The operation of each element in Fig. 2.2 is explained in the following sections.

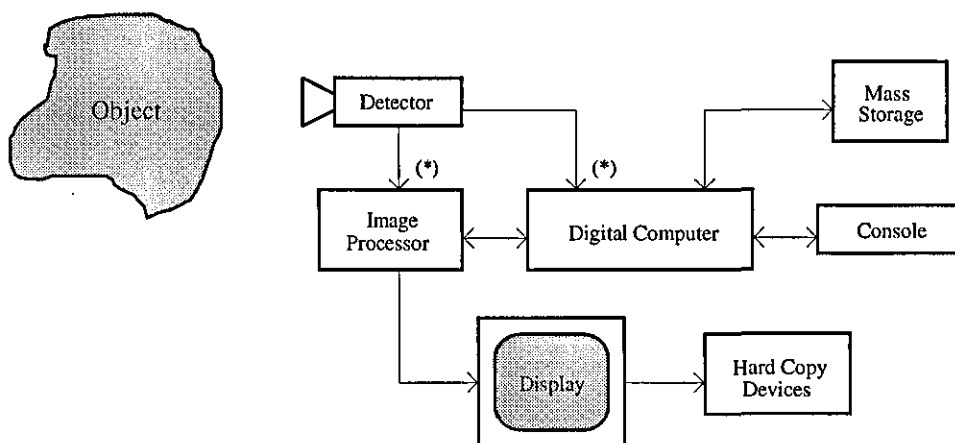


Fig. 2.2 - Elements of a digital image processing system.

(*) - Typically only one of two connections is made.

2.3.1 Image Detectors

Image detection and recording are divided into two major technologies: photochemical and photoelectric [4]. Photochemical technology is exemplified by the photographic film and photoelectric technology is exemplified by the television or video camera. Both technologies are widespread in digital image processing.

Photochemical methods have the advantage of combining image detection and recording into a single compact entity: the film. Photoelectric systems, on the other hand, usually require separation of the image detection process from the image recording process. However, photoelectric systems sense and convert images into an electric video signal that is ideally suitable for input into an image processor.

Image detection and recording by photographic film rely upon the properties of halide salts of silver. Silver halides, as AgCl and AgBr, are changed by exposure to light (the exact nature of the change is still not definitely known) so that the action of mild reducing agents, known as developers, results in the precipitation and deposition of grains of free silver. It has been found [5] that the mass of silver deposited is linearly proportional to the logarithm of the total exposure E , where

$$E = \int_{t_1}^{t_2} I(t) dt \quad (2.3-1)$$

and $I(t)$ is the optical intensity incident on the film. There is, in addition, a region of saturation, where all the available silver is deposited due to the intensity of the light, and a region of fog, where some small amount of silver is deposited even in the absence of any exposure. This behaviour of film is usually displayed in a curve of $D - \log E$, where D is the optical density, and is defined as

$$D = \log_{10} \left(\frac{I_i}{I_r} \right), \quad (2.3-2)$$

where I_i is the intensity of the incident light that illuminates a photographic film and I_r is the intensity of the light transmitted through or reflected from the film.

A typical $D - \log E$ curve is shown in Fig. 2.3 and is sometimes called the characteristic curve of a film [5] or the $H - D$ curve, after Hurter and Driffield who discovered it. To analyze the recording properties of a film, let an incident intensity of I_o illuminate the

film for an interval of time such that the total exposure E falls in the linear region of the curve in Fig. 2.3.

Hurter and Driffield showed [6] that the measured optical density D from Eq. (2.3-2) in the linear region of Fig. 2.3 can be expressed by

$$D = \gamma \log_{10} E - D_o, \quad (2.3-3)$$

where γ is the slope of the linear region, referred to as the gamma of the film, and D_o is the offset (the linear portion does not pass through the origin).

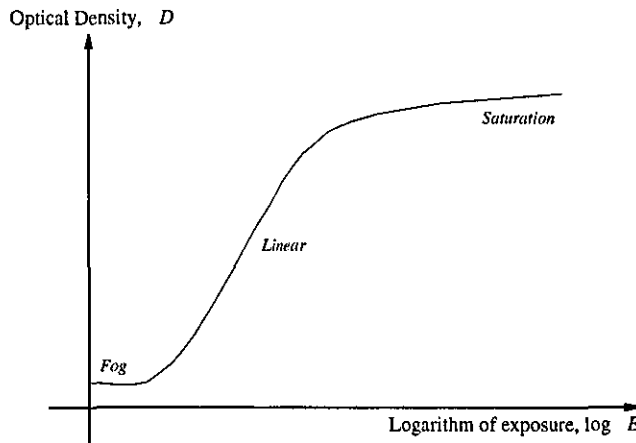


Fig. 2.3 - Typical $D - \log E$ curve of a film.

If it is assumed that the exposure was constant during a unit time interval, so that from Eq. (2.3-1) $E = I_o$, then substituting Eq. (2.3-3) into Eq. (2.3-2) the result is

$$I_r = I_i \exp(-\gamma \log_{10} I_o + D_o) = CI_o^{-\gamma}. \quad (2.3-4)$$

Equation (2.3-4) is the basic equation of film behaviour as a device that records image intensities. It is noted that:

- the intensity I_r observed is, in general, a highly non-linear mapping of the intensity that originally exposed the film, except for $\gamma = -1$ which corresponds to a linear mapping of the incident intensity;
- a film with positive γ is known as a negative material and a film with negative γ is known as a positive material;
- the slope of the curve γ determines the contrast characteristics of the photographic material.

Only in the case of a positive film of unit slope is found linear mapping of the incident intensity to observed intensity transmitted through the film. Thus, film is a non-linear element in the process of detecting and recording images.

Photoelectric image detection devices usually consist of three principal sub-systems: optics to form the image, the photo active surface to detect the image, and scanning electronics, to read out the electronic image formed on the photo active surface [7]. Among the most commonly used devices are vidicon cameras and photosensitive solid-state arrays.

Viewed as a system, the input or stimulus variable of a photoelectric image detector is radiant energy (visible or invisible light, X-ray photons, etc.). The output or response variable is an electrical quantity (current or voltage), typically the current in an electron beam scanning detector or a voltage in a solid-state detector. Response over a wide range of input and output variables is common, so it is conventional to plot the logarithm of the output variable as a function of the logarithm of the input variable.

As in the case of film, there is a saturation response at high intensities and at low intensities there is a limit response, usually referred to as dark-current (response of the detector with no incident radiation). There exists a linear region in which the logarithm of response is proportional to the logarithm of stimulus. This is analogous to film also, and it is conventional practice to refer to the slope of the linear portion of the curve as the gamma of the detector.

In vidicon cameras [1,4] the image is focused directly on the surface of a photo sensitive tube whose response is somehow proportional to the incident light pattern. The operation of vidicon cameras is based on the principle of photo conductivity. An independent, finely focused electron beam scans the rear surface of the photo conductive target and, by charge neutralization, this beam creates a potential difference that produces on a collector a current signal proportional to the brightness pattern.

Based on that, the behaviour in the linear region can be expressed in a form similar to Eq. (2.3-3):

$$\log i_b = \gamma \log I_o - \xi, \quad (2.3-5)$$

where i_b is the scanning beam current, I_o is the incident illumination intensity, γ is the linear slope, and ξ is the offset. Thus, the relation

$$i_b = KI_o^\gamma \quad (2.3-6)$$

is obtained for the transfer characteristics between input light intensity and the measured electron beam current.

There are two practical differences between the responses represented by the Eqs. (2.3-4) and (2.3-6). First, the photoelectric detector produces a response proportional to the incident light quanta; whereas photographic film has a negative exponent behaviour. Second, there are many photoelectric systems for which the input-output characteristics are linear over a wide range, that is $\gamma=1$, and deviations from linearity are not too severe even in the non-linear devices.

The solid-state arrays [8-10], also called charge coupled devices (CCD's), charge injection devices (CID's), and charge transfer devices (CTD's), are composed of discrete silicon imaging elements, called photosites, that have also an output signal proportional to the intensity of the incident light. Solid-state arrays are organized in one or two principal geometrical arrangements: line scan sensors and area sensors. A line scan sensor consists of a row of photosites and produces a two-dimensional image by relative motion between the scene and the detector. An area sensor is composed of a matrix of photosites and is therefore capable of capturing an image in the same manner as, say, a vidicon tube. Typically a high resolution area sensor has 576 lines of 768 elements on a $8.5 \times 8.5 \mu\text{m}$ grid.

Vidicon and area sensors are typically packaged as TV video cameras and the image digitization is achieved by feeding the output video signal of the camera into the acquisition module of an image processor as it will be discussed in the next section. Some detectors have the capability to perform the image digitization themselves and the digital output signal can be fed directly into a digital computer.

2.3.2 Image Processors

A digital image processor is the heart of any image processing system. An image processor consists of a set of hardware modules that perform four basic image functions: acquisition, storage, display and low-level (fast) processing [1-3,11].

Typically, the acquisition is performed by an acquisition module. The video input component in this module is, in general, an analogue video signal from a detector. First,

the video signal is processed and the video source is synchronized with the module in a suitable way. Then the signals are digitized, both spatially and in intensity (brightness), and stored after pre-processing, with the input look-up table, in a storage module.

Image processors which only process black-and-white video signals include a video multiplex input. This allows the acquisition module to be connected to a number of video signals. Via software, it is possible to switch between different video sources. Before the analogue video signal is digitized, it passes through a video amplifier with programmable gain and offset. In this way, the incoming video signal can best be adapted to the input range of the analogue-digital converter (ADC).

Colour acquisition modules can digitize colour images from RGB video sources that provide separate red, green, and blue video signals. Composite colour video signals in which luminance and colour information are composed in one video signal cannot be captured. It is necessary to use a decoder that splits the composite signal into its RGB components. Such a decoder is already built into some image processors. These devices can also decode Y/C video signals according to the new S-VHS video standard. For scientific applications, instead of RGB video signals, video signals from several synchronized black-and-white video cameras can be captured. In this way, multi-channel images can be taken and processed simultaneously.

Besides the image contents, a video signal contains synchronization signals that mark the beginning of an image (vertical synchronization) and the image rows (horizontal synchronization). These signals are extracted from the video input by a synchronization stripper and used to synchronize the storage and display modules.

Most modern image processors can process video signals according to both the European CCIR (PAL) and the American RS-170 (NTSC) norms [1-3] with 25 and 30 frames per second, respectively. The acquisition module is often referred to as a frame grabber because it can snap a single frame image.

Generally, the video signal is digitized with a resolution of 8 bits, that is, 256 grey levels at a rate of 10 million pixels per second (10 MHz). The resulting digitized image contains 512×512 and 480×512 pixels in the European and American formats, respectively. For a long time, this has been a standard. Modern frame grabbers, however, are much more flexible. Thus the number of rows and columns and the pixel clock rate can be programmed in a wide range.

Standard video cameras work in the interlaced mode. This means that a frame is composed of two half frames that either contain the even or odd lines. These half frames, also called fields, are scanned one after the other. Consequently, the two fields are illuminated one after the other, with a time shift of half the time interval between two images. A single image containing moving objects flickers because it is composed by two half fields illuminated at different times.

The digitized image is written into a storage module, often called frame buffer, via an input look-up table (ILUT) [3]. ILUT's allow a fast implementation of homogeneous point operations before the pixels being stored in the frame buffer. The ILUT's can be externally accessed through a digital computer for programming purposes.

All image processors contain a frame buffer that can hold pixels quantified with at least 8 bits (256 grey levels) of information. Usually, the frame buffer can hold 2 to 16 full resolution 512×512 images. This corresponds to 0.5 to 4 Mbytes of memory. Modular image processors have the advantage that the frame buffer memory can be expanded. Frame buffers show a complex inner structure that allows digitized images to be written and read out simultaneously. While a new pixel value is written into a memory cell, the old contents is simultaneously read out. Newer frame buffer architectures allow independent writing (digitization) and reading (display) of pixels into and from memory, respectively.

Besides the continuous and sequential access of the frame buffer, for example for image display, a digital computer can randomly access the frame buffer memory for image transfer between the image processor and the computer and vice versa. Conversely, the frame buffer can be addressed at TV rates by a display module. This module reads an image memory converting the stored digital information into an analogue video signal so that the image can be viewed on a TV monitor or stored on a video recording device. In general, the output of the display module is a three-channel video output, one channel for each colour red, green and blue. In this way, colour images can be generated even from grey level imagery. In a grey level frame buffer, the digitized value is sent out to all the channels simultaneously. Before the grey levels are converted to an analogue video signal by the digital-analogue converter (DAC), they pass by an output look-up table (OLUT) [3] for homogeneous pixel operations. Each colour channel has its own OLUT. If all the three OLUT's contain the same operation table, a grey image will be displayed. By programming this OLUT's with different tables, grey value images can be converted into pseudo colour images. In colour image processors each output channel is

connected to a different part of the frame buffer and true colour images, or more generally, multi-channel images with 8 bits per channel, can be displayed.

Basically all image processing still has to be performed on a digital computer. This also includes all the information required to transfer the image data from the frame buffer to the computer and back again. Because of the relatively slow transfer and processing speed on a computer, image processing operations cannot be performed in real time. To avoid this problem, a processing module is usually integrated on the image processor. This module includes dedicated hardware that allows common image processing operations to be speeded up considerably. There are two possibilities:

- special processing elements can be integrated into the frame buffer which are dedicated to a single type of operation;
- or a fast general purpose processor can be integrated.

Special processing elements have the advantage that certain operations can be performed in real time [2,11]. Of course, such a solution does not provide much flexibility and is not suitable for implementing complex algorithms. Such processing elements will predominantly be applied to low-level (fast) image processing tasks, while freely programmable processors will be more efficient for complex algorithms in later image processing stages. An arithmetic-logic unit (ALU) is normally integrated within the processing module. With this unit, arithmetic and logic operations can be performed in real time between two or more images.

2.3.3 Digital Computers

As mentioned in the previous section, an image processor may be equipped with internal processing capabilities, but the level of this processing is rather low in sophistication. Thus one usually finds that image processors are interfaced to a general-purpose digital computer, which provides versatility as well as ease of programming.

Computer systems used for image processing range from microprocessor devices to large computer systems capable of performing computationally intensive functions on large image arrays. The principal parameters influencing the structure of a computer for image processing are the intended application and the required data throughput. For dedicated applications (which normally dictate low cost) a well-equipped microcomputer or minicomputer will often be sufficient. If the application involves

extensive program development or is characterized by significant data throughputs, a workstation or a mainframe computer would most likely be required. In this case, a computer with virtual memory-addressing capabilities has significant advantages. Virtual addressing makes disk peripheral storage available to the user as if it were main memory. This feature, which is available in UNIX based operating systems [12], is of crucial importance because digital images utilize large amounts of memory during processing. The alternative to virtual memory-addressing is a set of user-supplied complex routines whose only function is to swap image segments in and out of the peripheral storage during processing. The disadvantage of this alternative is that the processing time is considerably increased due to the almost constant swapping of the image segments.

It should be mentioned that the standard high-level languages for computer programming such as C, FORTRAN, and PASCAL give great versatility and flexibility on the implementation of complex algorithms for image processing when compared with the non-standard low-level machine languages.

2.3.4 Storage Devices

A typical digital image consisting of 512×512 pixels, each of which is quantified into 8 bits, requires 0.25 Mbytes of storage. Thus providing adequate bulk storage facilities is one of the most important aspects in the design of a general-purpose image processing system. The three principal storage media used in this type of work are magnetic disks, magnetic tapes, and optical disks. Magnetic disks with a capacity of 400 Mbytes or more are common. A 400 Mbyte disk would hold on the order of 1,600 images of the size mentioned above. High-density magnetic tapes (6,400 bytes per inch) can store one of such image in approximately forty inches. Optical disks, which are based on laser writing and reading technology, have recently become commercially available. The storage capacity of a single optical disk platter can approach 4 Gbytes, which translates into approximately 16,000 images per disk.

2.3.5 Display and Recording Devices

Monochrome and colour television monitors are the principal display devices used in modern image processing systems. Monitors are driven by the video output of the display module in the image processor, as presented in Section 2.3.2. These signals can

also be fed into a video image recording device whose function is to produce a hard copy (slides, photographs, and transparencies) of the image being viewed on the monitor screen.

Other display media include cathode ray tubes (CRT's) and printing devices like laser printers, heat-sensitive printers, and ink-spray systems that are usually connected directly to a digital computer.

2.4 The Developed System

This section is devoted to the developed digital image processing system. The system is now fully operational in the Optoelectronics Laboratory at the Loughborough University of Technology, and follows the basic structure shown in Fig. 2.2.

Usually, a set of base constraints will be imposed by the applications that will be implemented with the system. Thus, the selection of the adequate equipment will be conditioned by them and it is very important that the selected equipment possessed enough flexibility and capabilities to accommodate further applications. In this case in particular, the base constraints were:

- image size up to 512×512 pixels with 256 grey levels.
- simultaneous image acquisition from up to 3 different monochrome video sources (PAL) synchronized together.
- triggered image acquisition by an external trigger signal.
- capability to run large iterative and interactive algorithms with large image arrays.

A detailed description of the hardware components and the software of the system is given in the next sections.

2.4.1 Hardware Description

The digital image processor used is a Real Time Colour Frame Grabber model S2200 [13]. It is a single width SBus card for the Sun Microsystems SPARCstation and compatible SBus based computers. It accepts composite video (PAL/NTSC), component analogue video RGB, or S-VHS video Y/C, and decodes this to form a 24 bit word of digital RGB data, which is stored in an on-board frame buffer memory in real time. The

stored image can be output continuously to a dedicated monitor along with a superimposed computer generated overlay. The frame buffer memory can be read back and modified by the computer, and displayed on the computer screen. No computer resources are needed while the image is being digitized unless the image is also displayed on the computer screen.

The board, in this particular case, is supplied with a PAL decoder, which has software to control and adjust the colour level. Either the output of this decoder or the RGB inputs drive the three 8 bit analogue-digital converters (ADC), one each for the red, green, and blue channels. Control of the converters by software allows the brightness, contrast, gain, and offset to be adjusted.

Input look-up tables (ILUT's) on each of the ADC outputs followed by matrix operations allow programmable mapping to be performed on the data from the ADC before it is written into the frame buffer memory, for instance to provide gamma correction, YCrCb to RGB conversion, and other arbitrary manipulations.

The frame buffer is a 1.75 Mbytes of dual ported video memory (VRAM). This is directly mapped into the computer memory space and can be read to and written from as a normal memory. Many storage modes are available, but the most commons are: one 576×768 image (24 bit colour) or three 576×768 images (8 bit monochrome).

A dedicated colour monitor can be connected to display in real time an image stored in video memory. This image is formed from 24 bits of captured RGB data (before or after computer processing), and 4 bits of non-destructive overlay data (for cursor, text, and so on), which are passed through output look-up tables (OLUT's) for each of the digital-analogue converters (DAC's). As the ILUT's, the OLUT's allow a programmable mapping to be performed between the data from the video memory and that displayed on the monitor. As standard this board accepts CCIR (PAL) standard interlaced video as input. For video output, non-interlaced 50 Hz frame rate video is also supported to give an improved quality picture with negligible flicker. Software enables the board to synchronize to input video or to generate its own video synchronization wave forms from its internal oscillator.

The board can operate in colour or monochrome modes. An important feature that can be achieved in colour mode is related with multi-channel monochrome acquisition. Up to three channels of monochrome video sources can be captured simultaneously. This means that the 8 bits of data from video source 1 appear in the red plane buffer, data

from video source 2 in the green plane buffer, and data from video source 3 in the blue plane buffer. Therefore reading one 24 bit word gives data from the same pixel location from up to three different sources. The video sources must be connected to the RGB inputs and they need to be synchronized together. The board has also a hardware feature that enables triggered acquisition by setting the appropriate signal into the external trigger input. This causes image acquisition to start on the next valid frame and stop after acquisition of one complete image.

As video sources the system uses two solid-state CCD cameras Pulnix model PE530. These are high-resolution monochrome cameras with composite video output signal in format CCIR (PAL). The active area on the sensors has 582 lines of 752 pixel elements on a $8.5 \times 8.5 \mu\text{m}$ grid with eight shutter speed settings (between 1/60 and 1/10,000 sec). Also, they have the capability to perform electronically gamma correction operation, as mentioned in Section 2.3.1, for two different gamma values ($\gamma = 1$ linear and $\gamma = 0.45$ non-linear). In multi-channel acquisition these cameras can be synchronized together, one camera being the source of the synchronism signals for the other. The gain control on these cameras can be set to automatic or manual which gives an excellent flexibility in particular applications.

The digital computer is a workstation SPARCclassic from Sun Microsystems [14,15] with the new processor MicroSPARC working at 50 MHz. The configuration consists of 24 Mbytes of main memory, 424 Mbytes of internal disk storage, 1.44 Mbytes floppy disk media and a 16 inch colour monitor with 8 bit plane and $1,152 \times 900$ resolution. An auxiliary colour video monitor Goldstar and a 64 grey level thermal printer are used as external display and record devices, respectively.

2.4.2 Software Description

This section describes the computer program called *ptfX* that has been specially developed and implemented for image processing purposes. It is an object oriented application that supports the hardware features of the S2200 frame grabber discussed in the previous section. The program is written in C language [16,17] and runs under X Windows environment, in particular, under OPEN LOOK user interface (UI) [18-21]. The software developments and debug [14,22] have been done on the workstation SPARCclassic running the UNIX SunOS Solaris 2.1 and SunC Compiler 2.0.1 mentioned in the previous section.

The main functions included in the program concern image acquisition control, management, processing, and analysis of image data. Some of them are dedicated to drive directly the S2200 frame grabber through the supplied device driver. This device driver provides a standardized interface between the S2200 hardware and the application program. The full source code listings, not included in this thesis, are available in the Optoelectronics Laboratory for use by developers in order to make easy any modification or integration of new functions.

The program is invoked by typing on the computer command line *ptfX* followed by an option. The option controls how the program will work. As an example, to invoke the program in a system with the S2200 hardware installed and a grey scale colour map, the command must be *ptfX -Sc0*. If the system does not have the S2200 installed, the command would be *ptfX -Nc0*. The command *ptfX* without any option results in a usage on-line help message.

The result of such commands is one base window, called the panning main window, with a number of options: Set-up, File, Acquisition, Display, Processing, and Quit. They are activated by the mouse pointer as a push button. Each option has associated with it a pop-down menu with functions that can be executed. Some functions are executed immediately after activation by the mouse pointer. However, other functions need to display a dialogue-box for setting parameters or specific options before its execution. All the dialogue-boxes, like the pop-down menus, are mouse driven. Notices and error messages will be displayed when a missing or incorrect parameter/option is detected during execution.

A brief description about each option in the panning main window is given below:

- **Set-up** - Used to perform a full S2200 hardware configuration and the system reset.
- **File** - Used to load, save, and print image data stored in the frame buffers or in image files. Three image file formats are supported. They are the TIFF format, SUN Raster format, and general binary formats all identified by the extensions ".tif", ".sun", ".gif", ".bif" and ".ptf", respectively.
- **Acquisition** - Used to start and stop live image data acquisition. The triggered acquisition is also controlled in this option. When image acquisition is performed a window for live display of the acquired image data is opened on the computer screen.
- **Display** - Used to open, close and re-size windows for 2D display image data stored in the frame buffers. Also, pseudo 3D display with four different views is available.

- **Processing** - Used to perform image data transference between the frame buffers with or without pan and scroll facilities. Basic functions as arithmetic and logic operations, rotation, zooming, scaling, sampling, normalizing, and thresholding image data can be executed under this option. Special pop-down menus are provided for image enhancement and edge detection. Different algorithms can be selected in order to achieve the best result. System calibration and measurements can be carried out as well as image data filtering, profiling and frame buffer initializations. Two pop-down menus are exclusively dedicated respectively for the two-dimensional Fourier transformations including the convolution and correlation operations, and image data reconstruction from its Fourier components.
- **Quit** - Used to abandon the program.

2.4.3 External Trigger and Synchronization

Many areas of research involve image acquisition and processing for different purposes. In some of them, it is essential to take images at a specific instant of time. This means that the digital image processing system must acquire the image data without losing any information. In general, experimental arrangements are the source of image events, and a push button is normally used to start the event. In order to acquire any moment of the extremely fast event it is essential to synchronize accurately and consistently, the event and imaging system.

With the external trigger facility available in the S2200 frame grabber, it will be possible to perform synchronized image acquisition. The push button information that starts the event will be used to trigger the frame grabber for acquisition. However, the correct use of such a feature implies a complete understanding of how the S2200 frame grabber works in the trigger mode. When the application program sets the S2200 hardware for external triggered acquisition, it starts frame acquisition continuously. If the external trigger input is asserted in this mode, a complete image is stored in the frame buffer on the next vertical sync signal. As an asynchronous event, this must be synchronized with the vertical syncs running into the frame grabber.

To achieve this, it was necessary to build an external trigger unit to drive the trigger input of the frame grabber as well as the event. The trigger unit performs the following tasks:

- receives and de-bounce the push button information (Fire Button);

- sends it back to trigger the frame grabber (through the external trigger input);
- waits for the next valid vertical sync and sends a signal to trigger the event.

The schematic circuit diagram of the trigger unit is shown in Fig. 2.4.

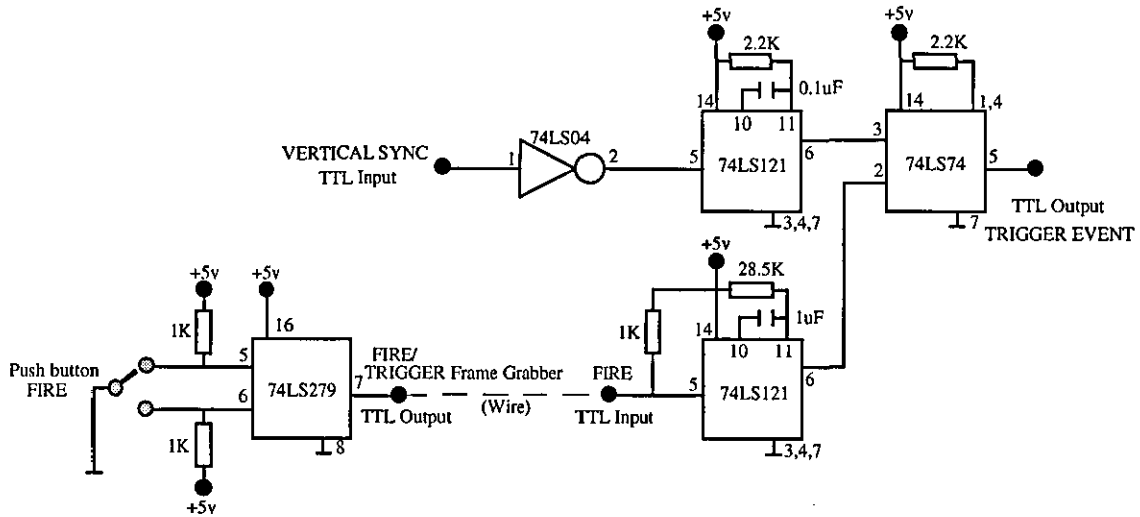


Fig 2.4 - Trigger Unit schematic circuit diagram.

With this trigger unit, synchronized image acquisition from an event will always be guaranteed. Delay units can be attached to this particular unit in order to introduce a variable time delay between the acquisition and event start.

2.5 Summary

In this chapter a complete digital image processing system has been presented and discussed. The system was physically implemented and it is currently available in the Optoelectronics Laboratory (Department of Physics) at the Loughborough University of Technology. As a powerful tool, this system has been used to support almost all the research carried out in this laboratory.

The digital image fundamentals, focusing image sampling and quantification, together with a detailed description of the individual elements in a basic general purpose image processing system have been used as a starting point for the development and later for the implementation of the system mentioned above. The system has been fully described and, in a general way, its capabilities and flexibility have also been discussed.

References

1. R. J. Schalkoff, *Digital Image Processing and Computer Vision*, John Wiley & Sons, Inc., New York (1989).
2. R. C. Gonzalez and P. Wintz, *Digital Image Processing*, 2nd Edition, Addison-Wesley Publishing Company, Reading, Massachusetts (1987).
3. B. Jähne, *Digital Image Processing: Concepts, Algorithms, and Scientific Applications*, 2nd Edition, Springer-Verlag, Berlin (1993).
4. H. C. Andrews and B. R. Hunt, *Digital Image Restoration*, Prentice-Hall, Inc. Englewood Cliffs, New Jersey (1977).
5. C. E. K. Mees and T. H. James, *The Theory of the Photographic Process*, Macmillan Company, New York (1966).
6. F. Hurter and V. C. Driffield, "Photo-Chemical Investigations and a New Method of Determination of the Sensitiveness of Photographic Plates", *J. Soc. Chemical Industry* **9**, 455-469 (1890).
7. L. M. Biderman and S. Nudelman, *Photoelectronic Imaging Devices*, Plenum Press, New York (1971).
8. D. Buss and R. Melen, *Charge-Coupled Devices: Technology and Applications*, IEEE Press, New York (1977).
9. D. F. Barbe, *Charge-Coupled Devices*, Springer-Verlag, Berlin (1980).
10. G. S. Hobson, *Charge-Transfer Devices*, Edward Arnold, London (1978).
11. W. K. Pratt, *Digital Image Processing*, John Wiley & Sons, Inc., New York (1978).
12. R. Morgan, H. McGilton, *Introducing UNIX™ System V*, McGraw-Hill, Inc., New York (1987).

13. *S2200 Real Time Colour Frame Grabber and Display - User's Manual*, Data Cell Limited and Active Silicon Limited, United Kingdom (1992).
14. *Desktop Sparc - System User's Guide*, Sun Microsystems Inc., California (1991).
15. *Desktop Sparc - Hardware Owner's Guide*, Sun Microsystems Inc., California (1992).
16. B. W. Kernighan and D. M. Ritchie, *The C Programming Language*, 2nd Edition, Prentice-Hall, Inc., Englewood Cliffs, New Jersey (1988).
17. C. A. Lindley, *Practical Image Processing in C*, John Wiley & Sons, Inc., New York (1991).
18. D. Heller, *XView Programming Manual*, 3rd Edition, O'Reilly & Associates, Inc., United States of America (1992).
19. T. V. Raalte, *XView Reference Manual*, O'Reilly & Associates, Inc., United States of America (1993).
20. A. Nye, *XLib Programming Manual*, O'Reilly & Associates, Inc., United States of America (1993).
21. A. Nye, *XLib Reference Manual*, 3rd Edition, O'Reilly & Associates, Inc., United States of America (1993).
22. *SparCompilers C 2.0.1 - Programmer's Guide*, Sun Microsystems, California (1992).

CHAPTER 3

Image Transformations

3.1 Introduction

Image transformation theory has played a key role in image processing for a number of years, and it continues to be topic of interest in theoretical as well as applied work in this field. Two-dimensional transformations are used for image enhancement, filtering, reconstruction, encoding, and description. In particular, this chapter deals with the development of two-dimensional Fourier transforms because of its wide range of applications in image processing problems.

The Fourier transformation of one and two continuous variables is introduced in Section 3.2, and then these concepts are expressed in discrete form in Section 3.3 together with some two-dimensional illustrative examples. The principal properties of the Fourier transformation are also presented and described in Section 3.4. A reference to the Fast Fourier transform algorithm is given in Section 3.5. Finally, Section 3.6 includes a brief summary.

3.2 Continuous Fourier Transformation

Let $f(x)$ be a continuous function of a real variable x . The Fourier transform of $f(x)$, denoted $FT\{f(x)\}$, is defined by the equation

$$FT\{f(x)\} = F(u) = \int_{-\infty}^{+\infty} f(x) \exp[-j2\pi xu] dx, \quad (3.2-1)$$

where $j = \sqrt{-1}$.

Given $F(u)$, $f(x)$ can be obtained by using the inverse Fourier transform

$$FT^{-1}\{F(u)\} = f(x) = \int_{-\infty}^{+\infty} F(u) \exp[j2\pi xu] du. \quad (3.2-2)$$

Equations (3.2-1) and (3.2-2), which are called Fourier transform pair, can be shown to exist if $f(x)$ is continuous and integrable and $F(u)$ is integrable [1-4]. These conditions are almost always satisfied in practice.

Throughout this thesis functions $f(x)$ that are real or complex valued will be considered. The Fourier transform of such functions is generally a complex valued function $F(u)$, that can be expressed by

$$F(u) = R_F(u) + jI_F(u), \quad (3.2-3)$$

where $R_F(u)$ and $I_F(u)$ are, respectively, its real and imaginary components. It is often convenient to express Eq. (3.2-3) in exponential form [4]:

$$F(u) = |F(u)| \exp[j\psi(u)], \quad (3.2-4)$$

where

$$|F(u)| = \sqrt{R_F^2(u) + I_F^2(u)} \quad (3.2-5)$$

and

$$\psi(u) = \tan^{-1} \left[\frac{I_F(u)}{R_F(u)} \right]. \quad (3.2-6)$$

The magnitude function $|F(u)|$ is called the Fourier spectrum or spectral magnitude of $f(x)$, and $\psi(u)$ its phase angle or spectral phase. The square of the spectrum,

$$P(u) = |F(u)|^2 = R_F^2(u) + I_F^2(u) \quad (3.2-7)$$

is commonly referred to as the power spectrum [4,5] of $f(x)$. The term spectral density is also commonly used to denote the power spectrum.

The variable u appearing in the Fourier transform is often called the frequency variable. This name arises from the fact that, using Euler's formula, the exponential term, $\exp[-j2\pi xu]$, may be expressed in the form:

$$\exp[-j2\pi xu] = \cos(2\pi xu) - j\sin(2\pi xu). \quad (3.2-8)$$

If the integral in Eq. (3.2-1) is interpreted as a limit-summation of discrete terms, it is evident that $F(u)$ is composed of an infinite sum of sine and cosine terms, and that each

value of u determines the frequency of its corresponding sine-cosine pair. Thus, this transformation enables the function $f(x)$ to be expressed and analysed in a different domain usually referred to as the spectral or frequency domain.

The Fourier transform can be easily extended to a function $f(x, y)$ of two variables. If $f(x, y)$ is continuous and integrable, and $F(u, v)$ is integrable [1-4], then the following Fourier pair exists:

$$FT\{f(x, y)\} = F(u, v) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) \exp[-j2\pi(xu + yv)] dx dy \quad (3.2-9)$$

and

$$FT^{-1}\{F(u, v)\} = f(x, y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} F(u, v) \exp[j2\pi(xu + yv)] du dv, \quad (3.2-10)$$

where u and v are the frequency variables.

As in the one-dimensional case, the spectral magnitude, spectral phase, and power spectrum are, respectively, given by the relations [4,5]:

$$|F(u, v)| = \sqrt{R_F^2(u, v) + I_F^2(u, v)}, \quad (3.2-11)$$

$$\psi(u, v) = \tan^{-1} \left[\frac{I_F(u, v)}{R_F(u, v)} \right], \quad (3.2-12)$$

and

$$P(u, v) = |F(u, v)|^2 = R_F^2(u, v) + I_F^2(u, v). \quad (3.2-13)$$

Using the Eqs. (3.2-11) and (3.2-12) the complex Fourier transformation can be expressed by the exponential notation

$$F(u, v) = |F(u, v)| \exp[j\psi(u, v)]. \quad (3.2-14)$$

The inverse Fourier transformation of $F(u, v)$, as it was mentioned in the beginning of this section, is a real or complex valued function $f(x, y)$ which is generally expressed by

$$f(x, y) = R_f(x, y) + jI_f(x, y), \quad (3.2-15)$$

where $R_f(x,y)$ and $I_f(x,y)$ are, respectively, its real and imaginary components. The Eq. (3.2-15) can also be expressed in the exponential form [4]:

$$f(x,y) = |f(x,y)| \exp[j\theta(x,y)], \quad (3.2-16)$$

where

$$|f(x,y)| = \sqrt{R_f^2(x,y) + I_f^2(x,y)} \quad (3.2-17)$$

and

$$\theta(x,y) = \tan^{-1} \left[\frac{I_f(x,y)}{R_f(x,y)} \right] \quad (3.2-18)$$

are, respectively, the spatial magnitude and spatial phase.

It should be noted that for real valued functions $f(x,y)$, the imaginary component $I_f(x,y)$ is zero and the spatial magnitude is defined by $|f(x,y)| = R_f(x,y)$ with constant spatial phase $\theta(x,y) = 2k\pi$, for $k = 0, \pm 1, \pm 2, \dots$

3.3 Discrete Fourier Transformation

Suppose a continuous function $f(x)$ is discretized into a sequence of values $\{f(x_0), f(x_0 + \Delta x), f(x_0 + 2\Delta x), \dots, f(x_0 + [N-1]\Delta x)\}$ by taking N samples Δx apart. It will be convenient in subsequent developments to use x as either a discrete or continuous variable, depending on the context of the discussion. This can be done by defining

$$f(x) = f(x_0 + x\Delta x), \quad (3.3-1)$$

where x now assumes the discrete values $0, 1, 2, \dots, [N-1]$. In other words, the sequence $\{f(0), f(1), f(2), \dots, f(N-1)\}$ will be used to denote any N uniformly spaced samples from a corresponding continuous function.

With the above notation in mind, the discrete Fourier transform pair that applies to sampled functions [4,6-8] is given by

$$F(u) = \frac{1}{N} \sum_{x=0}^{N-1} f(x) \exp[-j2\pi xu/N] \quad (3.3-2)$$

for $u = 0, 1, 2, \dots, [N-1]$, and

$$f(x) = \sum_{u=0}^{N-1} F(u) \exp[j2\pi xu/N] \quad (3.3-3)$$

for $x = 0, 1, 2, \dots, [N-1]$.

The values $u = 0, 1, 2, \dots, [N-1]$ in the discrete Fourier transform given in Eq. (3.3-2) correspond to samples of the continuous transform at values $0, \Delta u, 2\Delta u, \dots, [N-1]\Delta u$. In other words, $F(u)$ represents $F(u\Delta u)$. This notation is similar to that used for the discrete $f(x)$, with the exception that the samples of $F(u)$ start at the origin of the frequency axis. It can be shown [4,8] that Δu and Δx are related by the expression

$$\Delta u = \frac{1}{N\Delta x}. \quad (3.3-4)$$

In the two-variable case, the discrete Fourier transform pair is given by the equations [9]

$$F(u, v) = \frac{1}{NM} \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} f(x, y) \exp[-j2\pi(xu/N + yv/M)] \quad (3.3-5)$$

for $u = 0, 1, 2, \dots, [N-1]$; $v = 0, 1, 2, \dots, [M-1]$, and

$$f(x, y) = \sum_{u=0}^{N-1} \sum_{v=0}^{M-1} F(u, v) \exp[j2\pi(xu/N + yv/M)] \quad (3.3-6)$$

for $x = 0, 1, 2, \dots, [N-1]$ and $y = 0, 1, 2, \dots, [M-1]$.

Sampling of a continuous function is now in a two-dimensional grid with divisions of width Δx and Δy in the x and y axis. As in the one-dimensional case, the discrete function $f(x, y)$ represents samples of the function $f(x_0 + x\Delta x, y_0 + y\Delta y)$ for $x = 0, 1, 2, \dots, [N-1]$ and $y = 0, 1, 2, \dots, [M-1]$. Similar comments hold for $F(u, v)$. The sampling increments in the spatial and frequency domains are related by [4,8,9]

$$\Delta u = \frac{1}{N\Delta x} \quad (3.3-7)$$

and

$$\Delta v = \frac{1}{M\Delta y}. \quad (3.3-8)$$

When images are sampled in a square array then $M = N$ and

$$F(u, v) = \frac{1}{N} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x, y) \exp[-j2\pi(xu + yv)/N] \quad (3.3-9)$$

for $u, v = 0, 1, 2, \dots, [N-1]$, and

$$f(x, y) = \frac{1}{N} \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} F(u, v) \exp[j2\pi(xu + yv)/N] \quad (3.3-10)$$

for $x, y = 0, 1, 2, \dots, [N-1]$.

Note in this case that a $1/N$ term is included in both expressions. Since $F(u, v)$ and $f(x, y)$ are a Fourier transform pair, the grouping of these constant multiplicative terms is arbitrary. Images are typically digitized in square arrays, so the Fourier transform pair given in Eqs. (3.3-9) and (3.3-10) will be mostly considered in practice. The formulation given in Eqs. (3.3-5) and (3.3-6) will be used from time to time in situations where it is important to stress generality of the image size.

The Fourier magnitude, phase, and power spectrum of one- and two-dimensional discrete functions are also given by Eqs. (3.2-5) through (3.2-7) and Eqs. (3.2-11) through (3.2-13), respectively. The only difference is that the independent variables are discrete. For the two-dimensional discrete inverse Fourier transform the spatial magnitude and phase are also given by Eqs. (3.2-17) and (3.2-18) which are easily extended to the one-dimensional case.

Unlike the continuous case, it is not necessary to be concerned about the existence of the discrete Fourier transform because both $F(u)$ and $F(u, v)$ always exist in the discrete case [8].

As an example, Fig. 3.1 shows the two-dimensional discrete Fourier transform of a real valued function $f(x, y)$ for $N = M = 512$. The spatial magnitude function $|f(x, y)|$, shown in Fig. 3.1(a), is a centred rectangular aperture filled with a constant distribution of amplitudes and the spatial phase function $\theta(x, y)$, shown in Fig. 3.1(b), is constant and equal to zero. Figure 3.1(c) and (d) show, respectively, the Fourier magnitude and phase of the resulting $F(u, v)$.

Figure 3.2 shows another example of the two-dimensional discrete Fourier transform in which a spatial shifted real valued function is used instead. The spatial magnitude

function is the same one that has been used in the previous example but it was spatially shifted in both vertical and horizontal directions. The spatial phase function is constant and equal to zero.

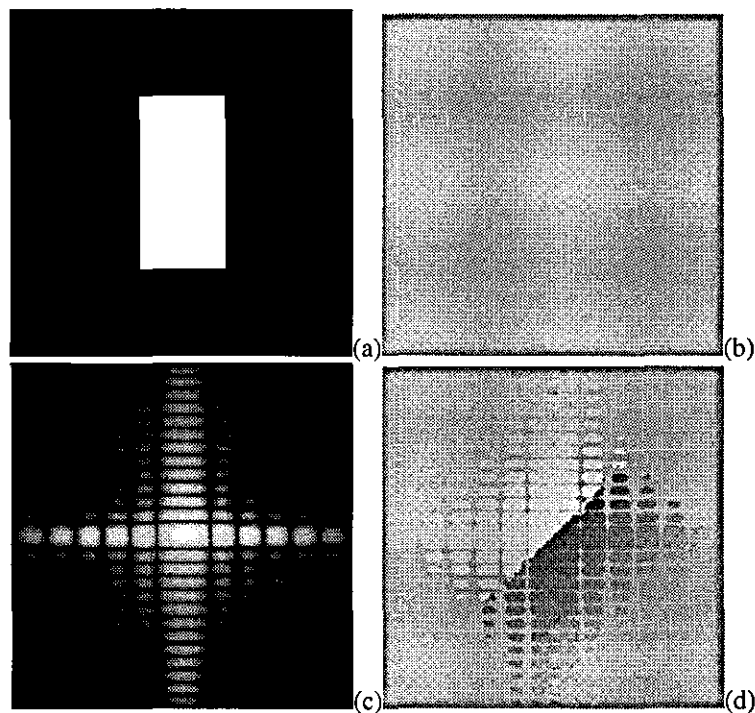


Fig. 3.1 - A two-dimensional real function and its Fourier transform. (a)- spatial magnitude; (b)- constant spatial phase; (c)- spectral magnitude; (d)- spectral phase.

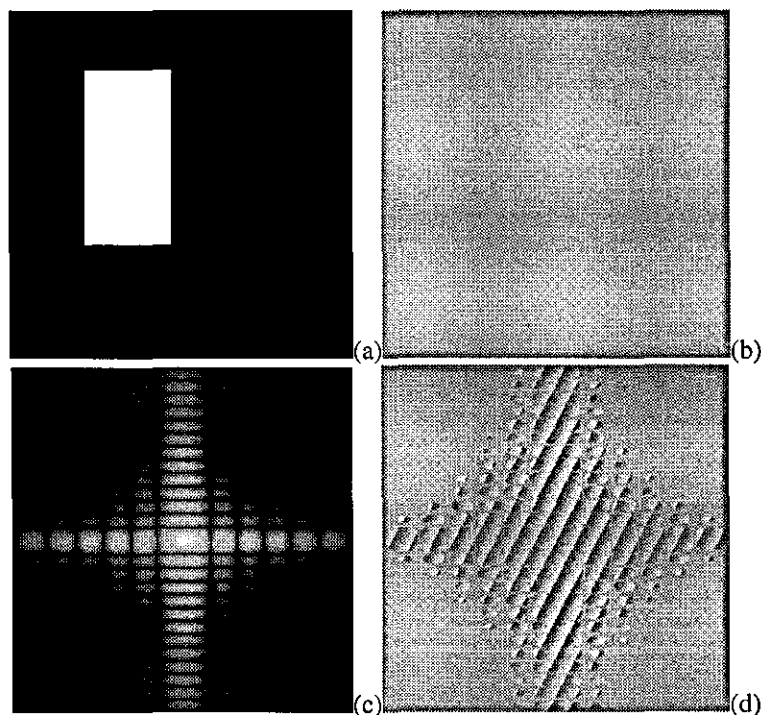


Fig. 3.2 - A two-dimensional shifted real function and its Fourier transform. (a)- spatial magnitude; (b)- constant spatial phase; (c)- spectral magnitude; (d)- spectral phase.

The results obtained for the Fourier magnitude and phase with this shifted real function are shown, respectively, in Fig. 3.2(c) and (d).

Analysing the results obtained for the Fourier magnitude, shown in Figs. 3.1(c) and 3.2(c), it can be observed that the modulus of the spectral magnitude functions are exactly the same but the spectral phase functions are quite different. The spectral phase change is totally related with the spatial shift [10] introduced by the real function used in the example shown in Fig. 3.2. This means that spectral magnitudes are not affected by spatial shifts but they considerably affect the spectral phases.

A last example, illustrated in Fig. 3.3, completes this comparison analysis, where a complex valued function is used instead. The spatial magnitude function, shown in Fig. 3.3(a), is the same as that used in the example shown in Fig. 3.1. The spatial phase function is the equivalent aperture filled with a linear phase distribution defined in the interval $[-\pi, +\pi]$ along the vertical direction, as shown in Fig. 3.3(b). Figure 3.3(c) and (d) show, respectively, the Fourier magnitude and phase of the resulting two-dimensional discrete Fourier transform.

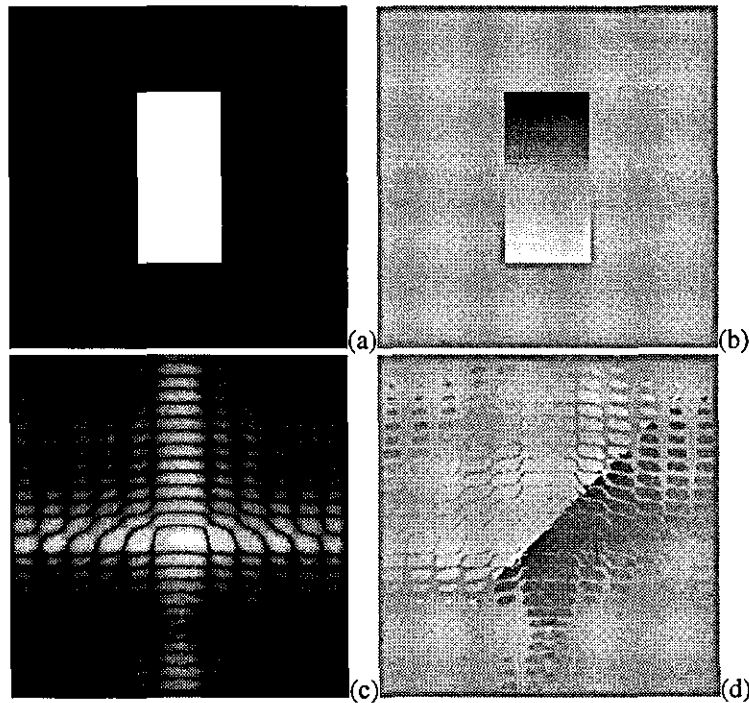


Fig. 3.3 - A two-dimensional complex function and its Fourier transform. (a)- spatial magnitude; (b)- linear spatial phase; (c)- spectral magnitude; (d)- spectral phase.

Comparing now the results obtained for the Fourier magnitude shown in Fig. 3.1(c) and 3.3(c), it can be observed that the spectral magnitude functions do not have the same modulus and one appears to be shifted with respect to the other in the frequency domain.

This frequency shift or spectral shift is totally related to the spatial phase change [10] introduced by the complex function used in the example shown in Fig. 3.3. The spectral phase functions, shown in Figs. 3.1(d) and 3.3(d), are now quite similar. This means that spectral magnitudes are considerably affected by linear spatial phase changes but they do not affect spectral phases. Similar comments hold for the case of non-linear spatial phase changes.

The purpose of these examples was to establish a comparison between the corresponding Fourier transform spectra in order to highlight some important characteristics of the Fourier analysis. For complex valued functions their Fourier spectra are always non-symmetric with respect to the origin of the frequency domain [10] (usually at the centre of the image). The information contained in the spectral phase concerns mostly the position in the spatial domain of individual elements in the original function, and the information about these individual elements is always contained in the spectral magnitude.

3.4 Fourier Transform Properties

Attention is focused in this section on properties of the Fourier transform that will be of value in subsequent discussions. Although our primary interest is in two-dimensional, discrete transforms, the underlying concepts of some of these properties are much easier to grasp if they are presented first in their one-dimensional, continuous form and then extended to the two-dimensional, discrete form.

Since several of the examples and results considered in this thesis are illustrated by images and their Fourier spectra displayed as intensity or magnitude functions, some comments concerning these displays are in order before beginning a discussion of Fourier transform properties. Many image spectra decrease rather rapidly as a function of increasing frequency [4,8] and therefore their high-frequency terms have a tendency to become obscured when displayed in image form. A useful processing technique that compensates for this difficulty consists of displaying the function

$$D(u, v) = \log(1 + |F(u, v)|) \quad (3.4-1)$$

instead of $|F(u, v)|$. This equation performs a logarithmic equalization of the values of the function $F(u, v)$ preserving the zero values in the frequency domain, since $D(u, v) = 0$ when $|F(u, v)| = 0$, resulting in an enhanced Fourier spectrum. It is also noted

that $D(u, v)$ is a non-negative function. Use of Eq. (3.4-1) in a two-dimensional display of the spectrum greatly facilitates the visualization and therefore the interpretation of the Fourier transformation. The equation was used throughout this thesis for enhancement purposes of all the Fourier spectrum displays shown.

3.4.1 Separability

The discrete Fourier transform pair given in Eqs. (3.3-9) and (3.3-10) can be expressed in the separable forms [11]

$$F(u, v) = \frac{1}{N} \sum_{x=0}^{N-1} \exp[-j2\pi xu/N] \sum_{y=0}^{N-1} f(x, y) \exp[-j2\pi yv/N] \quad (3.4-2)$$

for $u, v = 0, 1, 2, \dots, [N-1]$, and

$$f(x, y) = \frac{1}{N} \sum_{u=0}^{N-1} \exp[j2\pi xu/N] \sum_{v=0}^{N-1} F(u, v) \exp[j2\pi yv/N] \quad (3.4-3)$$

for $x, y = 0, 1, 2, \dots, [N-1]$.

The principal advantage of the separability property is that $F(u, v)$ and $f(x, y)$ can be obtained in two steps by successive applications of the one-dimensional Fourier transform or its inverse. This becomes evident if Eq. (3.4-2) is expressed in the form

$$F(u, v) = \frac{1}{N} \sum_{x=0}^{N-1} F(x, v) \exp[-j2\pi xu/N], \quad (3.4-4)$$

where

$$F(x, v) = N \left[\frac{1}{N} \sum_{y=0}^{N-1} f(x, y) \exp[-j2\pi yv/N] \right]. \quad (3.4-5)$$

For each value of x , the expression inside the brackets is a one-dimensional transform with frequency values $v = 0, 1, 2, \dots, [N-1]$. Therefore the two-dimensional function $F(x, v)$ is obtained by taking a transform along each row of $f(x, y)$ and multiplying the result by N . The desired result, $F(u, v)$, is then evaluated by taking a transform along each column of $F(x, v)$, as indicated by Eq. (3.4-4). The procedure is summarized in Fig. 3.4 which shows that this is very helpful when a two-dimensional Fourier transform algorithm is implemented.

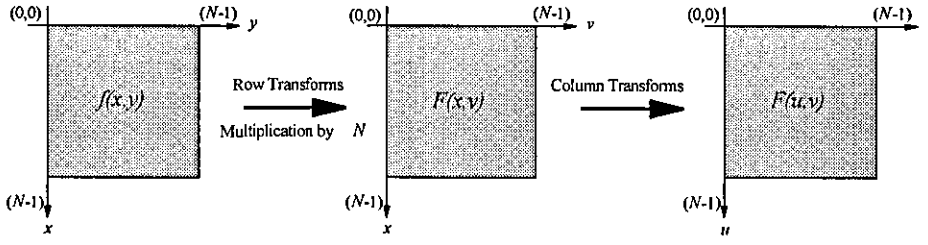


Fig. 3.4 - Computation of the two-dimensional Fourier transform as a series of one-dimensional transforms.

It should be noted that the same results would be obtained by first taking transforms along the columns of $f(x, y)$ and then along the rows of the result. Identical comments hold for the implementation of Eq.(3.4-3).

3.4.2 Translation

The translation properties of the Fourier transform pair are given by [10,11]

$$f(x, y) \exp[j2\pi(xu_0 + yv_0)/N] \Leftrightarrow F(u - u_0, v - v_0) \quad (3.4-6a)$$

and

$$f(x - x_0, y - y_0) \Leftrightarrow F(u, v) \exp[-j2\pi(x_0u + y_0v)/N], \quad (3.4-6b)$$

where the double arrow is used to indicate the correspondence between a function and its Fourier transform (and vice versa), as given in Eqs. (3.2-9) and (3.2-10) or Eqs. (3.3-9) and (3.3-10).

Equation (3.4-6a) shows that multiplying $f(x, y)$ by the indicated exponential term (spatial phase) and taking the Fourier transform of the product results in a shift of the origin of the frequency domain to the point (u_0, v_0) (spectral shift). Similarly, multiplying $F(u, v)$ by the exponential term (spectral phase) shown in Eq. (3.4-6b) and taking the inverse Fourier transform moves the origin of the spatial domain to (x_0, y_0) (spatial shift). This property was already outlined in the examples given in Section 3.3.

Equation (3.4-6a) with $u_0 = v_0 = N/2$ is often used in practical applications. In this case it follows that $\exp[j2\pi(xu_0 + yv_0)] = (-1)^{x+y}$ and

$$f(x, y)(-1)^{x+y} \Leftrightarrow F(u - N/2, v - N/2). \quad (3.4-7)$$

Thus the origin of the Fourier transform of $f(x,y)$, for convenience of displaying, can be moved to the centre of its corresponding $N \times N$ frequency square simply by multiplying $f(x,y)$ by $(-1)^{x+y}$ before the Fourier transformation.

It is noteworthy from Eq. (3.4-6a) that a spectral shift in $F(u,v)$ does not affect the modulus of the magnitude of its inverse Fourier transform, since

$$|f(x,y)\exp[j2\pi(xu_0 + yv_0)/N]| = |f(x,y)|. \quad (3.4-8)$$

This should be kept in mind especially when visual examination of the inverse Fourier transformation is limited to a display of its magnitude, being by necessity a display of its spatial phase, which contains the information concerning the spectral shift. Similarly, from Eq. (3.4-6b), a spatial shift in $f(x,y)$ does not affect the modulus of the magnitude of its Fourier transform, since

$$|F(u,v)\exp[-j2\pi(x_0u + y_0v)/N]| = |F(u,v)|, \quad (3.4-9)$$

being by necessity a display of its spectral phase when information concerning the spatial shift is required.

3.4.3 Periodicity and Conjugate Symmetry

The discrete Fourier transform and its inverse are periodic functions with period N , that is,

$$F(u,v) = F(u+N,v) = F(u,v+N) = F(u+N,v+N). \quad (3.4-10)$$

The validity of this property can be demonstrated by direct substitution of the variables $(u+N)$ and $(v+N)$ in Eq. (3.3-9). Although Eq. (3.4-10) points out that $F(u,v)$ repeats itself for an infinite number of values of u and v , only the N values of each variable in any one period are required to obtain $f(x,y)$ from $F(u,v)$. In other words, only one period of the transform is necessary to completely specify $F(u,v)$ in the frequency domain. For the particular case in which $f(x,y)$ is a real function, the Fourier transform also exhibits conjugate symmetry since,

$$F(u,v) = F^*(-u,-v) \quad (3.4-11)$$

or, more interestingly,

$$|F(u, v)| = |F(-u, -v)|. \quad (3.4-12)$$

This property plays an important role in the Fourier magnitude interpretation when it is displayed as an image.

3.4.4 Rotation

If the polar co-ordinates $x = r \cos \alpha$, $y = r \sin \alpha$, $u = \rho \cos \beta$ and $v = \rho \sin \beta$ are defined then $f(x, y)$ and $F(u, v)$ become $f(r, \alpha)$ and $F(\rho, \beta)$, respectively. It can be shown [12,13] by direct substitution in either the continuous or discrete Fourier transform pair that

$$f(r, \alpha + \alpha_0) \Leftrightarrow F(\rho, \beta + \alpha_0). \quad (3.4-13)$$

In other words, if $f(x, y)$ is rotated by an angle α_0 , then $F(u, v)$ is rotated by the same angle. Similarly, rotating $F(u, v)$ causes $f(x, y)$ to be rotated by the same angle.

3.4.5 Distributivity and Scaling

It follows directly from the definition of the continuous or discrete transform pair that

$$FT\{f_1(x, y) + f_2(x, y)\} = FT\{f_1(x, y)\} + FT\{f_2(x, y)\} \quad (3.4-14)$$

and, in general, that

$$FT\{f_1(x, y) \times f_2(x, y)\} \neq FT\{f_1(x, y)\} \times FT\{f_2(x, y)\}. \quad (3.4-15)$$

In other words, the Fourier transform and its inverse are distributive over the addition, but not over the multiplication.

It is also easy to show [4,13] that for the two scalars a and b ,

$$af(x, y) \Leftrightarrow aF(u, v) \quad (3.4-16)$$

and

$$f(ax, by) \Leftrightarrow \frac{1}{|ab|} F(u/a, v/b). \quad (3.4-17)$$

3.4.6 Average Value

A widely used definition of the average value of a two-dimensional discrete function is given by the expression [2,4,9]

$$\bar{f}(x, y) = \frac{1}{N^2} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x, y). \quad (3.4-18)$$

Substitution of $u = v = 0$ in Eq. (3.3-9) yields

$$F(0, 0) = \frac{1}{N} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x, y). \quad (3.4-19)$$

Therefore it is observed that $\bar{f}(x, y)$ is related to the Fourier transform of $f(x, y)$ by the equation

$$\bar{f}(x, y) = \frac{1}{N} F(0, 0). \quad (3.4-20)$$

3.4.7 Parseval-Rayleigh Theorem

The integral of the squared modulus of a function is equal to the integral of the squared modulus of its spectrum. This can be expressed for the two-dimensional continuous case by the equation

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |f(x, y)|^2 dx dy = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |F(u, v)|^2 du dv. \quad (3.4-21)$$

This theorem, which corresponds to Parseval's theorem for the Fourier series [1,2], was first used by Rayleigh [4] in his study of black-body radiation. The relation in Eq. (3.4-21) is based on the conservation of energy theorem in which each integral represents the amount of energy in a system. The theorem is sometimes referred to as the generalized Parseval-Rayleigh theorem and it can be expressed in the two-dimensional discrete case by the following relation

$$\sum_{x=0}^{N-1} \sum_{y=0}^{N-1} |f(x, y)|^2 = N^2 \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} |F(u, v)|^2. \quad (3.4-22)$$

3.4.8 Laplacian

The Laplacian of a two-variable function $f(x,y)$ is defined as

$$\nabla^2 f(x,y) = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}. \quad (3.4-23)$$

It follows from the definition [11,12] of the two-dimensional Fourier transform that

$$FT\{\nabla^2 f(x,y)\} \Leftrightarrow -(2\pi)^2(u^2 + v^2)F(u,v). \quad (3.4-24)$$

The Laplacian operator is useful for outlining edges in an image (edge detection).

3.4.9 Convolution and Correlation

This section considers two Fourier transform relationships that constitute a basic link between the spatial and frequency domains. These relationships, called convolution and correlation, are of fundamental importance in developing a firm understanding of image processing techniques based on the Fourier transformation. In order to clarify the concepts involved, this discussion begins by considering convolution in one dimension and with continuous arguments. The development is then extended to the discrete case and, finally, to the two-dimensional continuous and discrete cases. The same format is followed in developing the concept of correlation.

Convolution

The convolution of two functions $f(x)$ and $g(x)$, denoted by $f(x)*g(x)$, is defined by the integral [4,11-14]

$$f(x)*g(x) = \int_{-\infty}^{+\infty} f(\alpha)g(x-\alpha)d\alpha, \quad (3.4-25)$$

where α is a dummy variable of integration.

One aspect of Eq. (3.4-25), which will be of use later in this section, involves the convolution of a function $f(x)$ with the impulse function $\delta(x-x_0)$, which is defined by the relation

$$\int_{-\infty}^{+\infty} f(x)\delta(x-x_0)dx = f(x_0). \quad (3.4-26)$$

The function $\delta(x-x_0)$ may be viewed as having an area of unity in an infinitesimal neighbourhood about x_0 and being 0 everywhere else, that is,

$$\int_{-\infty}^{+\infty} \delta(x-x_0)dx = \int_{x_0^-}^{x_0^+} \delta(x-x_0)dx = 1. \quad (3.4-27)$$

For most purposes it can be said that $\delta(x-x_0)$ is located at $x=x_0$, and that the strength of the impulse is determined by the value of $f(x)$ at $x=x_0$. For instance, if $f(x)=A$, then $A\delta(x-x_0)$ is an impulse of strength A located at $x=x_0$.

As an example of the use of Eq. (3.4-25), suppose that the function $f(x)$ shown in Fig. 3.5(a) is convoluted with the function $g(x) = \delta(x+x_0) + \delta(x) + \delta(x-x_0)$ shown in Fig. 3.5(b). By folding $g(x)$, sliding it past $f(x)$, and making use of Eqs. (3.4-25) and (3.4-26), one obtains the result shown in Fig. 3.5(c). It is noted that convolution in this case amounts merely to copying $f(x)$ at the location of each impulse.

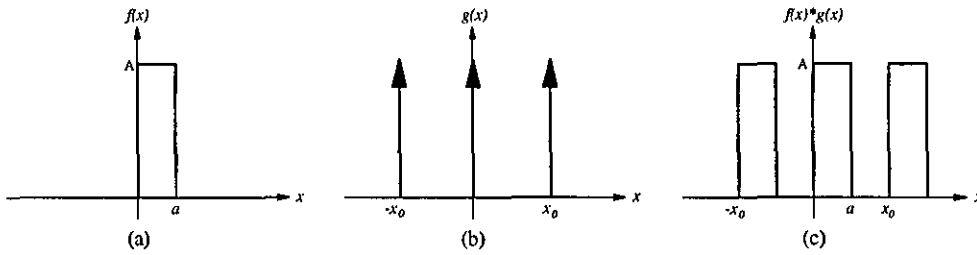


Fig. 3.5 - Convolution involving impulse functions.

The importance of convolution in frequency domain analysis lies in the fact that $f(x)*g(x)$ and $F(u)G(u)$ constitute a Fourier transform pair. In other words, if $f(x)$ has the Fourier transform $F(u)$ and $g(x)$ has the Fourier transform $G(u)$, then $f(x)*g(x)$ has the Fourier transform $F(u)G(u)$. This result, formally stated as

$$f(x)*g(x) \Leftrightarrow F(u)G(u) \quad (3.4-28)$$

indicates that convolution in the space domain can also be obtained by taking the inverse Fourier transform of the product $F(u)G(u)$. An analogous result is that convolution in the frequency domain reduces to multiplication in the space domain, that is,

$$f(x)g(x) \Leftrightarrow F(u)*G(u). \quad (3.4-29)$$

These two results are commonly referred to as the convolution theorem [11,13].

Suppose now that instead of being continuous, functions $f(x)$ and $g(x)$ are discretized into sampled arrays of size A and B , respectively: $\{f(0), f(1), f(2), \dots, f(A-1)\}$, and $\{g(0), g(1), g(2), \dots, g(B-1)\}$. As pointed out in Section 3.4.3, the discrete Fourier transform and its inverse are periodic functions. In order to formulate a discrete convolution theorem that is consistent with this periodicity property, it may be assumed that the discrete functions $f(x)$ and $g(x)$ are periodic with the same period M . The resulting convolution will then be periodic with the same period. The problem is how to select a value for M . It can be shown [11] that unless it is chosen

$$M \geq A + B - 1 \quad (3.4-30)$$

the individual periods of the convolution will overlap. This overlap is commonly referred to as the wraparound error. If $M = A + B - 1$, the periods will be adjacent and if $M > A + B - 1$, the periods will be spaced apart, with the degree of separation being equal to the difference between M and $A + B - 1$. Since the assumed period must be greater than either A or B , the length of the sampled sequences must be increased so that both are of length M . This can be done by appending zeros to the given samples to form the following extended sequences:

$$f_e(x) = \begin{cases} f(x) & 0 \leq x \leq A-1 \\ 0 & A \leq x \leq M-1 \end{cases}$$

and

$$g_e(x) = \begin{cases} g(x) & 0 \leq x \leq B-1 \\ 0 & B \leq x \leq M-1 \end{cases}$$

Based on this, the discrete convolution of $f_e(x)$ and $g_e(x)$ is defined by the expression

$$f_e(x) * g_e(x) = \sum_{m=0}^{M-1} f_e(m) g_e(x-m) \quad (3.4-31)$$

for $x = 0, 1, 2, \dots, [M-1]$. The convolution function is a discrete, periodic array of length M , with the values $x = 0, 1, 2, \dots, [M-1]$ describing a full period of $f_e(x) * g_e(x)$.

The mechanics of discrete convolution are basically the same as for continuous convolution. The only differences are that displacements take place in discrete increments corresponding to the separation between samples, and that integration is replaced by a summation. Similarly, Eqs. (3.4-28) and (3.4-29) also hold in the discrete

case where to avoid wraparound error extended functions $f_e(x)$ and $g_e(x)$ and their transforms $F_e(u)$ and $G_e(u)$ should be used. The discrete variables x and u assume values in the range $0, 1, 2, \dots, [M-1]$.

The two-dimensional convolution is analogous in form to Eq. (3.4-25). Thus for two functions $f(x, y)$ and $g(x, y)$, it is defined as

$$f(x, y) * g(x, y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(\alpha, \beta) g(x - \alpha, y - \beta) d\alpha d\beta. \quad (3.4-32)$$

The convolution theorem in two dimensions [11, 13], then, is given by the relations

$$f(x, y) * g(x, y) \Leftrightarrow F(u, v) G(u, v) \quad (3.4-33)$$

and

$$f(x, y) g(x, y) \Leftrightarrow F(u, v) * G(u, v). \quad (3.4-34)$$

The two-dimensional, discrete convolution is formulated by letting $f(x, y)$ and $g(x, y)$ be discrete arrays of size $A \times B$ and $C \times D$, respectively. As in the one-dimensional case, these arrays must be assumed periodic with the same period M and N , in the x and y directions, respectively. Wraparound error in the individual convolution periods is avoided by choosing

$$M \geq A + C - 1 \quad (3.4-35)$$

and

$$N \geq B + D - 1. \quad (3.4-36)$$

The periodic sequences are formed by extending $f(x, y)$ and $g(x, y)$ as follows:

$$f_e(x, y) = \begin{cases} f(x, y) & 0 \leq x \leq A-1 \quad \text{and} \quad 0 \leq y \leq B-1 \\ 0 & A \leq x \leq M-1 \quad \text{or} \quad B \leq y \leq N-1 \end{cases}$$

and

$$g_e(x, y) = \begin{cases} g(x, y) & 0 \leq x \leq C-1 \quad \text{and} \quad 0 \leq y \leq D-1 \\ 0 & C \leq x \leq M-1 \quad \text{or} \quad D \leq y \leq N-1 \end{cases}$$

The two-dimensional convolution of $f_e(x, y)$ and $g_e(x, y)$ is given by the relation

$$f_e(x, y) * g_e(x, y) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f_e(m, n) g_e(x - m, y - n) \quad (3.4-37)$$

for $x = 0, 1, 2, \dots, [M-1]$ and $y = 0, 1, 2, \dots, [N-1]$. The $M \times N$ array given by this equation is one period of the discrete, two-dimensional convolution. If M and N are chosen according to the Eqs. (3.4-35) and (3.3-36), this array is guaranteed to be free of interference from other adjacent periods. As in the one-dimensional case, the continuous convolution theorem given in Eqs. (3.4-33) and (3.4-34) also apply to the discrete case with $u = 0, 1, 2, \dots, [M-1]$ and $v = 0, 1, 2, \dots, [N-1]$. All computations involve the extended functions $f_e(x, y)$ and $g_e(x, y)$, and their transforms $F_e(u, v)$ and $G_e(u, v)$.

The theoretical power of the convolution theorem will become evident in Section 3.4.10 where the sampling theorem is discussed. From a practical point of view, it is often more efficient to compute the discrete convolution in the frequency domain instead of using Eq. (3.4-37) directly. The procedure is to compute the Fourier transforms of the extended functions, then, the two transforms are multiplied and the inverse Fourier transform of the product will yield the convolution function.

Correlation

The correlation of two continuous functions $f(x)$ and $g(x)$, denoted by $f(x) \circ g(x)$, is defined by the relation [4,11-14]

$$f(x) \circ g(x) = \int_{-\infty}^{+\infty} f^*(\alpha) g(x + \alpha) d\alpha, \quad (3.4-38)$$

where $*$ means the complex conjugate function. If $f(x)$ and $g(x)$ are the same function, Eq. (3.4-38) is usually called the auto-correlation function or the cross-correlation function. The forms of Eqs. (3.4-25) and (3.4-38) are similar, the only difference being that the function $g(x)$ is not folded about the origin. Thus correlation is performed by simply sliding $g(x)$ by $f(x)$ and integrating the product from $-\infty$ to $+\infty$ for each value of displacement x .

The discrete equivalent of Eq. (3.4-38) is defined as

$$f_e(x) \circ g_e(x) = \sum_{m=0}^{M-1} f_e^*(m) g_e(x + m) \quad (3.4-39)$$

for $x = 0, 1, 2, \dots, [M-1]$. The comments made earlier regarding $f_e(x)$ and $g_e(x)$, the assumed periodicity of these functions, and the choice of values for M , also apply to Eq. (3.4-39).

Similar expressions hold for the two-dimensional case. Thus if $f(x,y)$ and $g(x,y)$ are functions of continuous variables, their correlation is defined as

$$f(x,y) \circ g(x,y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f^*(\alpha,\beta) g(x+\alpha, y+\beta) d\alpha d\beta. \quad (3.4-40)$$

In the discrete case, then

$$f_e(x,y) \circ g_e(x,y) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f_e^*(m,n) g_e(x+m, y+n) \quad (3.4-41)$$

for $x = 0, 1, 2, \dots, [M-1]$ and $y = 0, 1, 2, \dots, [N-1]$. As in the case of discrete convolution, $f_e(x,y)$ and $g_e(x,y)$ are extended functions, and M and N are chosen according to Eqs. (3.4-35) and (3.4-36) in order to avoid wraparound error in the periods of the correlation function.

It can be shown for both the continuous and discrete cases that the following correlation theorem holds [11,13]:

$$f(x,y) \circ g(x,y) \Leftrightarrow F^*(u,v) G(u,v) \quad (3.4-42)$$

and

$$f^*(x,y) g(x,y) \Leftrightarrow F(u,v) \circ G(u,v). \quad (3.4-43)$$

As in the case of discrete convolution, the computation of the discrete correlation of $f_e(x,y) \circ g_e(x,y)$ is often more efficiently carried out in the frequency domain by multiplying the Fourier transforms $F_e^*(u,v)$ and $G_e(u,v)$ of the extended functions $f_e(x,y)$ and $g_e(x,y)$, and finally inverting the Fourier transformation.

When comparing the results of discrete versus continuous convolution or correlation, it should be noted that because the way it has been defined, the discrete cases amount to evaluation of the continuous forms by rectangular integration. Thus if one wishes to compare discrete and continuous results on the absolute basis, Eqs. (3.4-31) and (3.4-39) should be multiplied by Δx , and Eqs. (3.4-37) and (3.4-41) by $\Delta x \Delta y$, where Δx and Δy are the separation between samples, as defined in Section 3.3. When one is only computing and evaluating the discrete forms, however, the inclusion of these scale factors is a matter of preference. It is also important to note that all convolution and correlation expressions hold if the considered functions f and g , along with their corresponding transforms functions F and G , are interchanged.

3.4.10 Sampling Theorem

The basic idea of image sampling [11-16] was introduced in chapter two Section 2.2.2 on an intuitive basis. The Fourier transform and the convolution theorem provide the tools for a deeper analytical study of this problem. In particular, it is of interest in looking at the question of how many samples should be taken so that no information is lost in the sampling process. Expressed differently, the problem is one of establishing under what sampling conditions a continuous image can be fully recovered from a set of sampled values. This analysis begins with the one-dimensional case and then the concepts will be extended to the two-dimensional case.

One-dimensional Functions

Consider the function $f(x)$ shown in Fig. 3.6(a), which is assumed to extend from $-\infty$ to $+\infty$, and suppose that the Fourier transform of $f(x)$ vanishes for values of u outside the interval $[-W, +W]$. The transform might appear as shown in Fig. 3.6(b). A function whose transform has this property for any finite value of W is called a band-limited function. It must be mentioned that the Fourier transform is a complex function and, for simplicity, in the following graphical illustrations only the magnitude of the transforms is shown. The ordinate axis, however, will be labelled with $F(u)$, $G(u)$, etc., to indicate that the concepts involved are valid for the complete transform, and not just its magnitude.

A sampled version of $f(x)$ can be obtained by simply multiplication of this function by a sampling function $s(x)$, which consists of a train of impulses Δx units apart, as shown in Fig. 3.6(c). Since, by the convolution theorem, multiplication in the space domain is equivalent to convolution in the frequency domain, then the Fourier transform shown in Fig. 3.6(f) is obtained for the product $s(x)f(x)$. It is noted that the transform is periodic with period $1/\Delta x$, and that the individual repetitions of $F(u)$ can overlap. In the first period, for example, the centre of the overlapped region will occur at $u = 1/(2\Delta x)$ if the quantity $1/(2\Delta x)$ is less than W . Therefore this problem can be avoided by selecting the sampling interval Δx so that $1/(2\Delta x) \geq W$ or

$$\Delta x \leq \frac{1}{2W}. \quad (3.4-44)$$

The result of decreasing Δx is shown in Fig. 3.6(g) and (h). The net effect is to separate the periods so that no overlap occurs. The importance of this operation lies in the fact that if one multiplies the transform of Fig. 3.6(h) by the function

$$G(u)=\begin{cases} 1 & -W \leq u \leq +W \\ 0 & \text{elsewhere} \end{cases} \tag{3.4-45}$$

it become possible to isolate $F(u)$ completely, as shown in Fig. 3.6(k). The inverse Fourier transform then yields the original continuous function $f(x)$. The fact that a band-limited function can be recovered completely from samples whose spacing satisfies Eq. (3.4-44) is known as the Whittaker-Shannon sampling theorem [17,18].

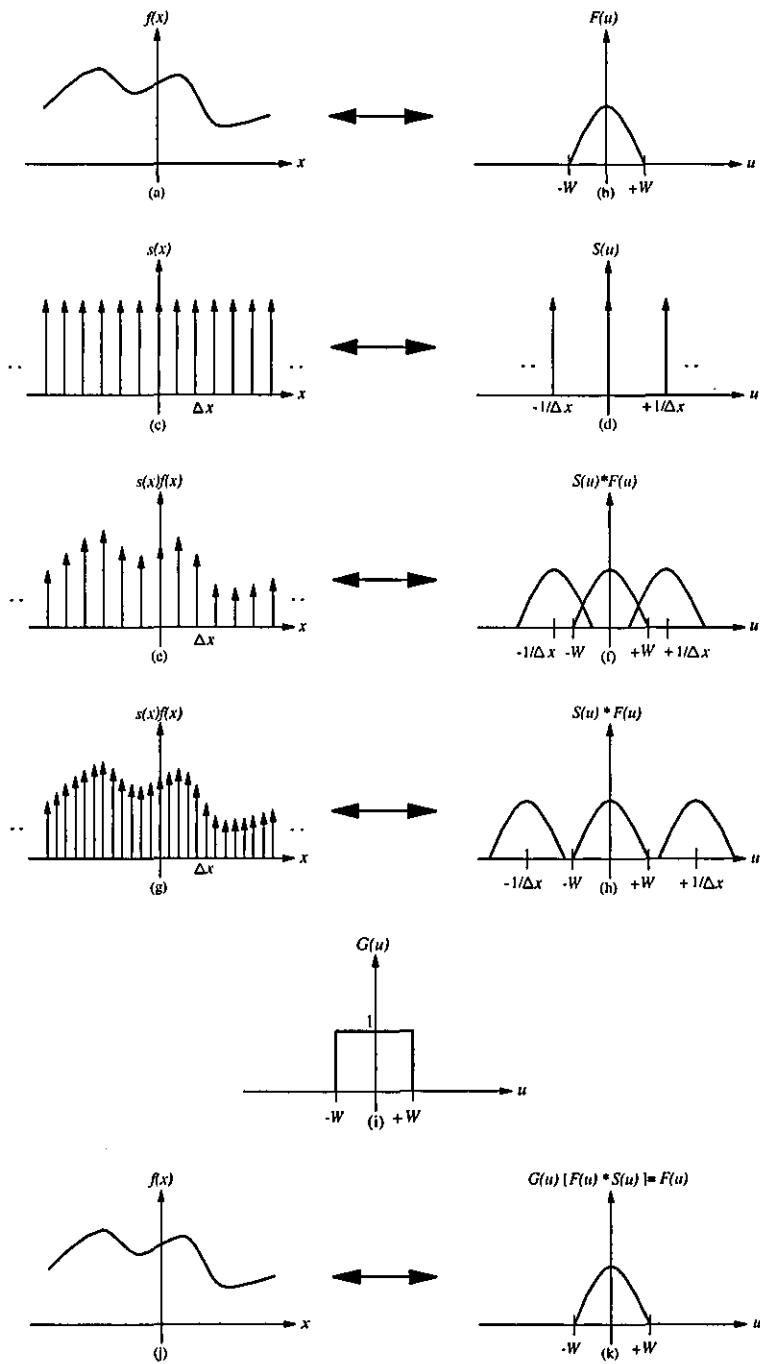


Fig. 3.6 - Graphical development of sampling concepts.

It is important to keep in mind that all the frequency domain information of a band-limited function is contained in the interval $[-W, +W]$. If Eq. (3.4-44) is not satisfied, however, the transform in this interval is corrupted by contributions from adjacent periods. This phenomenon, frequently referred to as aliasing, precludes complete recovery of an under sampled function.

The foregoing results apply to functions that are of unlimited duration in the space domain. Since this implies an infinite sampling interval, it is of interest to examine the practical case where a function is sampled only over a finite region. This situation is shown graphically in Fig. 3.7. Parts (a) through (f) are the same as in Fig. 3.6, with the exception that the separation between samples is assumed to satisfy the sampling theorem so that no aliasing is present.

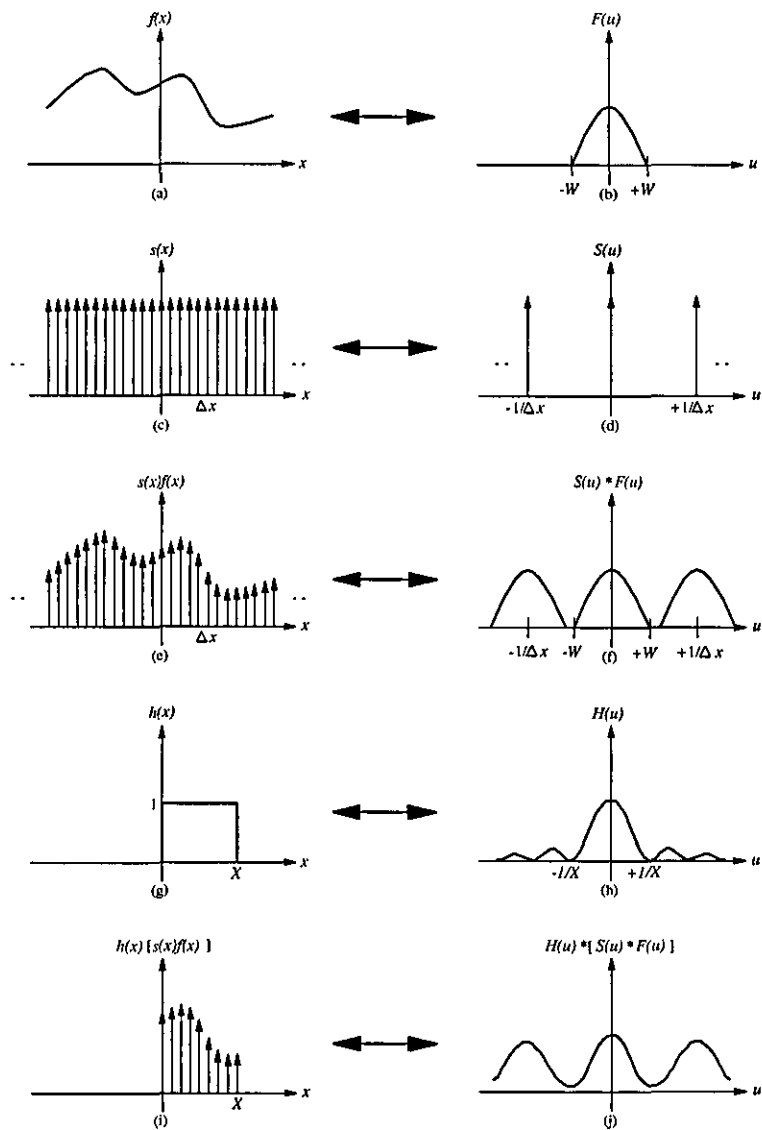


Fig. 3.7 - Graphical illustration of finite sampling concepts.

A finite sampling interval $[0, X]$ can be represented mathematically by multiplying the sampled result shown in Fig. 3.7(e) by the function

$$h(x) = \begin{cases} 1 & 0 \leq x \leq X \\ 0 & \text{elsewhere} \end{cases} \quad (3.4-46)$$

This function, often called a window, and its Fourier transform are shown in Fig. 3.7(g) and (h), respectively. The results of multiplication are illustrated in Fig. 3.7(i) and (j). It is important to note that the final frequency domain result is obtained by convoluting the function $S(u)*F(u)$ with $H(u)$, which is the Fourier transform of the window function $h(x)$. Since $H(u)$ has frequency components that extend to infinity, the convolution of these functions introduces a distortion in the frequency domain representation of a function that has been sampled and limited to a finite region by $h(x)$, as shown in Fig. 3.7(j). Thus even if the separation between samples satisfies the sampling theorem, it is generally impossible to recover completely a function that has been sampled only in a finite region of the space domain. This fact can be appreciated by noting that it becomes impossible under these conditions to isolate the original Fourier transform. The only exception to this is when $f(x)$ is band-limited and periodic with period equal to X . In this case it can be shown that the corruption due to $H(u)$ cancel out, thus allowing complete recovery of $f(x)$ if the sampling theorem is satisfied. It is important to note that the recovered function still extends from $-\infty$ to $+\infty$ and is non-zero outside the range in which $h(x)$ is zero. These considerations lead us to the important conclusion that no function $f(x)$ of finite duration can be band-limited. Conversely, a function that is band-limited must extend from $-\infty$ to $+\infty$ in the space domain. This is an important practical result because it establishes a fundamental limitation in the treatment of digital functions.

It is of interest to use the above result to give an alternate reason for the periodicity of the discrete Fourier transform. It should be noted that, thus far, all results in the frequency domain have been of a continuous nature. A discrete Fourier transform can be obtained by simply sampling the continuous transform with a train of impulses that are Δu units apart. The situation is depicted in Fig. 3.8, where use has been made of Fig. 3.7(i) and (j) with notation changed to $f(x)$ and $F(u)$ to facilitate comparison with the discussion in Section 3.3. It should be kept in mind, however, that these functions are assumed to be the result of the sequence of operations shown in Fig. 3.7.

As previously pointed out, sampling can be represented by multiplying the impulse train and the function of interest. In this case, $F(u)$ is multiplied by $S(u)$ and the result obtained is shown in Fig. 3.8(f). The equivalent operation in the space domain is

convolution, which yields the function shown in Fig. 3.8(e). This function is periodic with period $1/\Delta u$. If N samples of $f(x)$ and $F(u)$ are taken and the spacing between samples are selected so that a period in each domain is covered by N uniformly spaced samples, then $N\Delta x = X$ in the space domain and $N\Delta u = 1/\Delta x$ in the frequency domain.

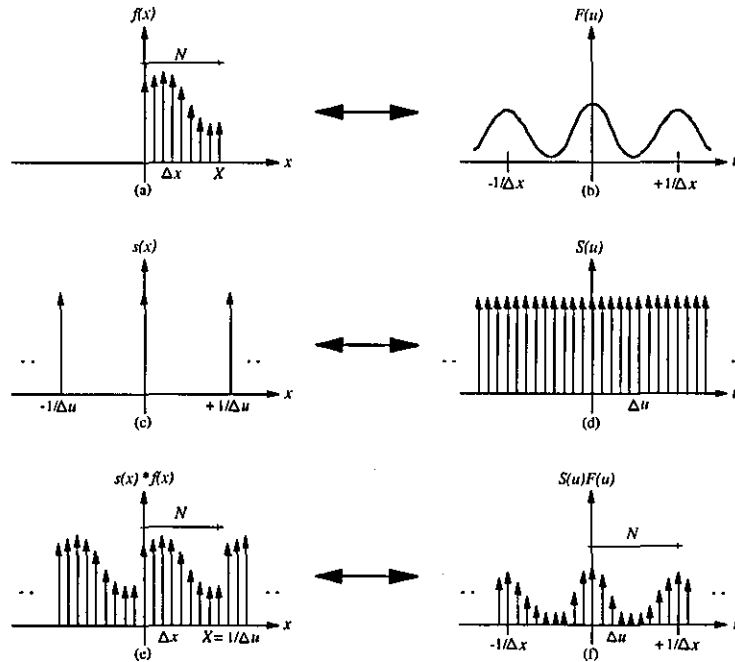


Fig. 3.8 - Graphical illustration of discrete Fourier transform.

The latter equation follows from the fact that the Fourier transform of a sampled function is periodic with period $1/\Delta x$, as shown earlier. It follows from this equation that

$$\Delta u = \frac{1}{N\Delta x} \quad (3.4-47)$$

which agrees with Eq. (3.3-4). Choosing this spacing yields the function in Fig. 3.8(e), which is periodic with period $1/\Delta u$. From Eq. (3.4-47) then $1/\Delta u = N\Delta x = X$ which is the overall sampling interval in Fig. 3.8(a).

Two-dimensional Functions

The sampling concepts developed above are, after some modifications in notation, directly applicable to two-dimensional functions. The sampling process for these functions can be formulated mathematically by making use of the two-dimensional impulse function $\delta(x, y)$ that is defined as follows:

$$\iint_{-\infty}^{+\infty} f(x,y)\delta(x-x_0,y-y_0)dx dy = f(x_0,y_0). \quad (3.4-48)$$

The interpretation of Eq. (3.4-48) is analogous to that given in connection with Eqs (3.4-26) and (3.4-27). So, the two-dimensional sampling function $s(x,y)$ consists of a train of impulses separated by Δx units in the x direction and Δy units in the y direction.

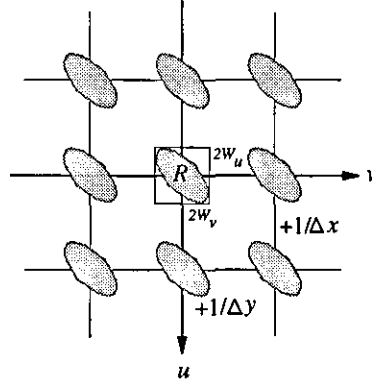


Fig. 3.9 - Frequency domain representation of a sampled two-dimensional, band-limited function.

Given the function $f(x,y)$, where x and y are continuous, a sampled function is obtained by forming the product $s(x,y)f(x,y)$. The equivalent operation in the frequency domain is convolution of $S(u,v)$ and $F(u,v)$, where $S(u,v)$ is a train of impulses with separation $1/\Delta x$ and $1/\Delta y$ in the u and v directions, respectively. If $f(x,y)$ is band-limited (i.e., its Fourier transform vanishes outside some finite region R) the result of convoluting $S(u,v)$ and $F(u,v)$ might look like the case shown in Fig. 3.9. It is noted that the function shown is periodic in two directions.

Let $2W_u$ and $2W_v$ represent the widths in the u and v directions, respectively, of the smallest rectangle that completely encloses the region R . Then, from Fig. 3.9, if $1/\Delta x \geq 2W_u$ and $1/\Delta y \geq 2W_v$, no aliasing is present and one of the periods can be recovered completely if the convolution function $S(u,v)*F(u,v)$ is multiplied by

$$G(u,v) = \begin{cases} 1 & (u,v) \text{ inside one of rect including } R \\ 0 & \text{elsewhere} \end{cases} \quad (3.4-49)$$

The inverse Fourier transform of $G(u,v)[S(u,v)*F(u,v)]$ yields $f(x,y)$.

The foregoing considerations lead to a form of the two-dimensional sampling theorem, which states that a band-limited function $f(x,y)$ can be recovered completely from samples whose separation is given by

$$\Delta x \leq \frac{1}{2W_u} \quad (3.4-50a)$$

and

$$\Delta y \leq \frac{1}{2W_v}. \quad (3.4-50b)$$

When $f(x,y)$ is spatially limited by using a two-dimensional window function $h(x,y)$ analogous to the function $h(x)$ used in Fig. 3.7, then the problem that the transform of the sampled function is distorted by the convolution of $H(u,v)$ and $S(u,v)*F(u,v)$ is again present. This distortion, which is due to the spatially limited nature of digital images, precludes complete recovery of $f(x,y)$ from its samples. As in the one-dimensional case, periodic functions are an exception, but images satisfying this condition are rarely found in practice.

An argument similar to the one developed for the one-dimensional case can be carried out to show how periodicity arises in the two-dimensional, discrete Fourier transform. For an $N \times M$ image, this analysis would also yield the following results:

$$\Delta u = \frac{1}{N\Delta x} \quad (3.4-51a)$$

and

$$\Delta v = \frac{1}{M\Delta y} \quad (3.4-51b)$$

which agree with Eqs. (3.3-7) and (3.3-8).

These relationships between the sampled separations guarantee that a complete two-dimensional period will be recovered by $N \times M$ uniformly spaced values in both the spatial and frequency domains.

3.5 Fast Fourier Transform

The fast Fourier transform is a method for efficiently computing the finite discrete Fourier transform of a real or complex function (discrete samples) defined in space or time domain [19,20]. This method of computing the discrete Fourier transform suddenly became widely known and soon revolutionized many fields where heavy computing was an impediment to progress.

The number of complex multiplication and additions required to implement directly Eq. (3.3-2) is proportional to N^2 . This can be seen easily by noting that, for each of the N values of u , expansion of the summation sign requires N complex multiplication of $f(x)$ by $\exp[-j2\pi xu/N]$ and $(N-1)$ additions of the results. The terms of $\exp[-j2\pi xu/N]$ can be computed once and stored in a table for all subsequent applications. For this reason, the multiplication of x by u in these terms is usually not considered a direct part of the implementation. In this way, the amount of time required to evaluate the complete Fourier transform is large, and for large values of N such implementation is impracticable.

In this section it is shown that, by properly decomposing Eq. (3.3-2), the number of multiply and add operations can be made proportional to $N \log_2 N$ [7,20,21]. The decomposition procedure is called the fast Fourier transform (FFT) algorithm. The reduction in proportionality from N^2 to $N \log_2 N$ multiply/add operations represents a significant saving in computational effort, as shown in Table 3.1.

N	N^2 (Direct FT)	$N \log_2 N$ (FFT)	$N / \log_2 N$ (Advantage)
2	4	2	2.00
4	16	8	2.00
8	64	24	2.67
16	256	64	4.00
32	1,024	160	6.40
64	4,096	384	10.67
128	16,384	896	18.29
256	65,536	2,048	32.00
512	262,144	4,608	56.89

Table 3.1 - A comparison N^2 versus $N \log_2 N$.

It is evident from the Table 3.1 that the fast Fourier transform approach offers a considerable computational advantage over a direct implementation of the Fourier transform particularly when N is relatively large. For example, a direct implementation of a transform for $N = 512 \times 512$ (a typical digital image) requires in order of 38 hours on a machine such as a Sun SPARCclassic. By contrast, the same job can be done in the same machine in about 9 seconds using a fast Fourier transform algorithm.

Attention is focused below on the development of a fast Fourier transform algorithm of one variable. As pointed out in Section 3.4.1, a two-dimensional Fourier transform can be readily computed by a series of applications of the one-dimensional transform.

3.5.1 FFT Algorithm

The fast Fourier transform algorithm developed in this section is based on the so called successive doubling method [21-24]. It will be convenient in the following discussion to express Eq. (3.3-2) in the form

$$F(u) = \frac{1}{N} \sum_{x=0}^{N-1} f(x) W_N^{xu}, \quad (3.5-1)$$

where

$$W_N = \exp[-j2\pi/N], \quad (3.5-2)$$

and N is assumed to be of the form

$$N = 2^n, \quad (3.5-3)$$

where n is a positive integer. Based on this, N can be expressed as

$$N = 2M, \quad (3.5-4)$$

where M is also a positive integer. Substitution of Eq. (3.5-4) into Eq. (3.5-1) yields

$$\begin{aligned} F(u) &= \frac{1}{2M} \sum_{x=0}^{2M-1} f(x) W_{2M}^{xu} \\ &= \frac{1}{2} \left\{ \frac{1}{M} \sum_{x=0}^{M-1} f(2x) W_{2M}^{(2x)u} + \frac{1}{M} \sum_{x=0}^{M-1} f(2x+1) W_{2M}^{(2x+1)u} \right\}. \end{aligned} \quad (3.5-5)$$

Since, from Eq. (3.5-2), $W_{2M}^{2xu} = W_M^{xu}$, Eq. (3.5-5) may be expressed in the form

$$F(u) = \frac{1}{2} \left\{ \frac{1}{M} \sum_{x=0}^{M-1} f(2x) W_M^{xu} + \frac{1}{M} \sum_{x=0}^{M-1} f(2x+1) W_M^{xu} W_{2M}^u \right\}. \quad (3.5-6)$$

Defining the following functions

$$F_{\text{even}}(u) = \frac{1}{M} \sum_{x=0}^{M-1} f(2x) W_M^{xu} \quad (3.5-7)$$

for $u = 0, 1, 2, \dots, [M-1]$, and

$$F_{odd}(u) = \frac{1}{M} \sum_{x=0}^{M-1} f(2x+1) W_M^{xu} \quad (3.5-8)$$

for $u = 0, 1, 2, \dots, [M-1]$, Eq. (3.5-6) then becomes

$$F(u) = \frac{1}{2} \{ F_{even}(u) + F_{odd}(u) W_{2M}^u \}. \quad (3.5-9)$$

Also, since $W_M^{u+M} = W_M^u$ and $W_{2M}^{u+M} = -W_{2M}^u$, it follows from Eqs. (3.5-7) through (3.5-9) that

$$F(u+M) = \frac{1}{2} \{ F_{even}(u) - F_{odd}(u) W_{2M}^u \}. \quad (3.5-10)$$

Careful analysis of Eqs. (3.5-7) through (3.5-10) reveals some interesting properties of these expressions. It is noted that a N -point transform can be computed by dividing the original expression into two parts, as indicated in Eqs. (3.5-9) and (3.5-10). Computation of the first half of $F(u)$ requires the evaluation of the two $(N/2)$ -point transforms given in Eqs. (3.5-7) and (3.5-8). The resulting values of $F_{even}(u)$ and $F_{odd}(u)$ are then substituted into Eq. (3.5-9) to obtain $F(u)$ for $u = 0, 1, 2, \dots, [N/2-1]$. The other half then follows directly from Eq. (3.5-10) without additional transform evaluations for $u = N/2, \dots, [N-1]$.

Implementation of Eqs. (3.5-7) through (3.5-10) constitute the successive doubling FFT algorithm. This name arises from the fact that a two-point transform is computed from two one-point transforms, a four-point transform from two two-point transforms, and so on, for any N that is equal to an integer power of 2.

3.5.2 Inverse FFT

Thus far, little has been said concerning the inverse Fourier transform. It turns out that any algorithm for implementing the discrete forward transform can also be used (with minor modifications in the input) to compute the inverse transform. To see this, let us consider Eqs. (3.3-2) and (3.3-3), which are repeated below:

$$F(u) = \frac{1}{N} \sum_{x=0}^{N-1} f(x) \exp[-j2\pi xu/N] \quad (3.5-11)$$

and

$$f(x) = \sum_{u=0}^{N-1} F(u) \exp[j2\pi xu/N] \quad (3.5-12)$$

Taking the complex conjugate of Eq. (3.5-12) and dividing both sides by N yields

$$\frac{1}{N} f^*(x) = \frac{1}{N} \sum_{u=0}^{N-1} F^*(u) \exp[-j2\pi xu/N]. \quad (3.5-13)$$

By comparing this result with Eq. (3.5-11) it is clear that the right side of Eq. (3.5-13) is in the form of the forward Fourier transform. Thus if $F^*(u)$ is used as input into an algorithm designed to compute the forward transform the result will be the quantity $f^*(x)/N$. Taking the complex conjugate and multiplying by N yields the desired inverse transformation. For two-dimensional square arrays the procedure is to take the complex conjugate of Eq (3.3-10), that is,

$$f^*(x, y) = \frac{1}{N} \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} F^*(u, v) \exp[-j2\pi(xu + yv)/N], \quad (3.5-14)$$

which is also in the form of the two-dimensional forward Fourier transform given in Eq. (3.3-9). It therefore follows that if $F^*(u, v)$ is used as input into an algorithm designed to compute the forward transform the result will be $f^*(x, y)$. By taking the complex conjugate of this result the function $f(x, y)$ will be obtained. In the case where $f(x)$ or $f(x, y)$ is real, the complex conjugate operation is unnecessary since $f(x) = f^*(x)$ and $f(x, y) = f^*(x, y)$ for real functions.

The fact that the two-dimensional transform is usually computed by successive passes of the one-dimensional transform is a frequent source of confusion when using the above technique to obtain the inverse. In other words, when using the one-dimensional algorithm to compute the two-dimensional inverse, the method is not to compute the complex conjugate after each row or column is processed. Instead, the function $F^*(u, v)$ is treated as if it were $f(x, y)$ in the forward, two-dimensional transform procedure summarized in Fig. 3.4 in Section 3.4.1. The complex conjugate (if necessary) of the result will yield the proper inverse.

3.6 Summary

The principal purpose of this chapter has been to present a theoretical foundation of image transformations and their properties. The emphasis placed on the Fourier

transform reflects its wide scope of applications in image processing with the corresponding computational implications. Within this framework, the essential points necessary for a basic understanding of these concepts have been developed and illustrated.

Important characteristics of the Fourier analysis concerning real and complex valued functions and their spectrum have been highlighted in a comparison study also presented, which will be of great value in subsequent discussions in the following chapters. The sampling theorem has also been considered among the conditions under which an image can be fully recovered from a set of sampled values.

The material on the fast Fourier transform is of particular interest because it constitutes the main tool that will be used throughout this thesis. The fast Fourier transform algorithm has been implemented and it was included on the application program developed and discussed in the previous chapter.

References

1. E. C. Titchmarsh, *Introduction to the Theory of Fourier Integrals*, Oxford University Press, New York (1948).
2. A. Papoulis, *The Fourier Integral and its Applications*, McGraw-Hill Ed., New York (1962).
3. A. Papoulis, *Systems and Transforms with Applications in Optics*, R. E. Krieger Publishing Company, Inc., Malabar, Florida (1981).
4. R. N. Bracewell, *The Fourier Transform and its Applications*, 2nd Edition, McGraw-Hill Ed., New York (1986).
5. R. B. Blackman and J. W. Tukey, *The Measurement of the Power Spectra*, Dover Publications, New York (1958).
6. J. W. Cooley, P. A. W. Lewis, and P. D. Welch, "Application of the Fast Fourier Transform to Computation of Fourier Integrals", *IEEE Trans. Audio and Electroacoustics* **AU-15**, 79-84 (1967).
7. J. W. Cooley, P. A. W. Lewis, and P. D. Welch, "The Fast Fourier Transform and its Applications", *IEEE Trans. Educ.* **E-12**, 27-34 (1969).
8. E. O. Brigham, *The Fast Fourier Transform*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey (1974).
9. J. S. Lim, *Two-dimensional Signal and Image Processing*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey (1990).
10. S. L. Marple, *Digital Spectral Analysis with Applications*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey (1987).
11. R. C. Gonzalez and P. Wintz, *Digital Image Processing*, Addison-Wesley Publishing Company, Reading, Massachusetts (1987).
12. B. Jähne, *Digital Image Processing: Concepts, Algorithms and Scientific Applications*, 2nd Edition, Springer-Verlag, Berlin (1993).

13. W. K. Pratt, *Digital Image Processing*, John Wiley & Sons, Inc., New York (1978).
14. A. V. Oppenheim, R. W. Schaffer, *Digital Signal Processing*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey (1975).
15. R. H. T. Bates and M. J. McDonnell, *Image Restoration and Reconstruction*, Clarendon Press, Oxford (1986).
16. R. J. Schalkoff, *Digital Image Processing and Computer Vision*, John Wiley & Sons, Inc., New York (1989).
17. E. T. Whittaker, "On the Functions which are Represented by the Expansions of the Interpolation Theory", *Proc. Royal Society of Edinburgh* **35**, 181-194 (1915).
18. C. E. Shannon, "Communications in the Presence of Noise", *Proc. IRE* **37**, 10-21 (1949).
19. J. W. Cooley, P. A. W. Lewis, and P. D. Welch, "Historical Notes on the Fast Fourier Transform", *IEEE Trans. Audio and Electroacoustics* **AU-15**, 76-79 (1967).
20. W. T. Cochran, J. W. Cooley, *et. al.*, "What is the Fourier Transform?", *IEEE Trans. Audio and Electroacoustics* **AU-15**, 45-55 (1967).
21. J. W. Cooley and J. W. Tukey, "An Algorithm for the Machine Calculation of Complex Fourier Series", *Math. Comput.* **19**, 297-301 (1965).
22. R. E. Blahut, *Fast Algorithms for Digital Signal Processing*, Addison-Wesley Publishing Company, Reading, Massachusetts (1985).
23. S. Winograd, "On Computing the Discrete Fourier Transform", *Math. Comput.* **32**, 175-199 (1978).
24. R. D. Preuss, "Very Fast Computation of the Radix-2 Discrete Fourier Transform", *IEEE Trans. Acoustics, Speech, and Signal Processing* **ASSP-30**, 595-607 (1982).

CHAPTER 4

Image Formation and Reconstruction

4.1 Interference and Diffraction

When two sets of waves are made to cross each other, for example the waves created by dropping two stones simultaneously in a quiet pool, interesting and complicated effects can be observed. In the region of crossing there are places where the disturbance is practically zero and others where it is greater than that given by either wave alone. A very simple law can be used to explain these effects, which states that the resultant displacement of any point is merely the sum of the displacements due to each wave separately. This is known as the principle of superposition and was first clearly stated by Thomas Young in 1802 [1,2]. The truth of this principle is at once evident by observation that after the waves have passed out of the region of crossing, they appear to have been entirely unmodified by the other set of waves. Magnitude, frequency, and all the other characteristics are just as if they had crossed an undisturbed space. This could hold only provided the principle of superposition was true. Two different observers can see different objects through the same aperture with clearness, whereas the light reaching the two observers had crossed in going through the aperture. The principle is therefore applicable with great precision to light, and one can use it in investigating the disturbance in regions where two or more light waves are superimposed.

It was stated above that two beams of light can be made to cross each other without either one producing any modification of the other after it passes beyond the region of crossing. In this sense the two beams do not interfere with each other. However, in the region of crossing, where both beams are acting at once, one can expect from the considerations in Section 1.4 that the resultant amplitude and intensity may be very different from the sum of those contributed by the two beams acting separately. This modification of intensity obtained by the superposition of two or more beams of light is called interference. If the resultant intensity is zero or in general less than expected from the separate intensities, one can have destructive interference, while if it is greater, one can have constructive interference. The phenomenon in its simpler aspects is rather difficult to observe, because of the very short wavelength of the light, and therefore was not recognized as such before 1800, when the corpuscular theory of light was predominant. Thomas Young was the first man to successfully demonstrate the interference of light and its wave character. In order to understand his crucial

experiment, the application to light of an important principle which holds for any type of wave motion should be considered first.

4.1.1 Huygens' Principle

When waves pass through an aperture or past the edge of an obstacle, they always spread to some extent into the region which is not directly exposed to the oncoming waves. This phenomenon is called diffraction [3]. In order to explain this bending of light, Huygens, nearly three centuries ago, proposed the rule that each point on a wave-front may be regarded as a new source of spherical waves. This principle has very far-reaching applications and will be used later in discussing the diffraction of light.

Figure 4.1 shows schematically the diffraction of waves passing through a small aperture. A set of plane waves approaches the barrier AB from the left, and the barrier contains an opening s of width somewhat smaller than the wavelength λ . At all points except s the waves will be either reflected or absorbed, but s will be free to produce a disturbance behind the screen. It is found experimentally, in agreement with the above principle, that the waves spread out from s in the form of semicircles.

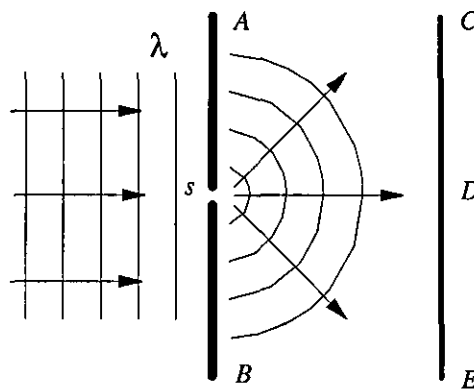


Fig. 4.1 - Diffraction of waves passing through a small aperture.

Huygens' principle as shown in Fig. 4.1 can be illustrated very successfully with water waves. If the experiment is performed with light, one would naturally expect, from the fact that light generally travels in straight lines, that merely a narrow patch of light would appear at D . However, if the slit is made very narrow, an appreciable broadening of this patch is observed, its breadth increasing as the slit is narrowed further. This is remarkable evidence that light does not always travel in straight lines and the waves on passing through a narrow opening spread out into a continuous fan of light rays. Observation of the light intensity distribution on the screen CE reveals that the light is

most intense in the forward direction, but its intensity decreases slowly as the angle increases. If the slit is small compared with the wavelength, the intensity does not come to zero even when the angle of observation becomes 90° .

4.1.2 Young's Experiment

The original experiment performed by Young is shown schematically in Fig. 4.2. Sunlight was first allowed to pass through a pinhole s and then, at a considerable distance away, through two pinholes s_1 and s_2 . The two sets of spherical waves emerging from the two holes interfered with each other in such a way as to form a symmetrical pattern of varying intensity on the screen AC .

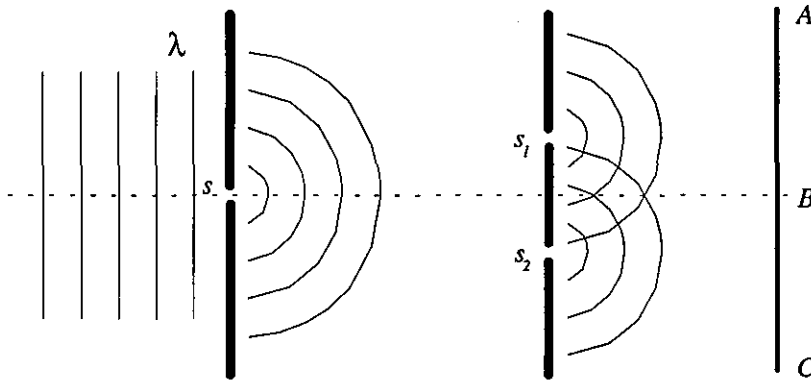


Fig. 4.2 - Experimental arrangement for Young's double-slit experiment.

Since this early experiment was performed, it has been found convenient to replace the pinholes by narrow slits and to use a source giving monochromatic light, that is, light of a single wavelength. If the circular lines represent crests of waves, the intersections of any two lines represent the arrival at those points of two waves with the same phase or with phases differing by a multiple of 2π . Such points are therefore those of maximum disturbance or brightness. A close examination of the light intensity distribution on the screen will reveal evenly spaced light and dark bands also called fringes.

One can now derive an equation for the intensity at any point P on the screen and investigate the spacing of the interference fringes. As shown in Fig. 4.3, two waves arrive at P having traversed different distances s_2P and s_1P . Hence they are superimposed with a phase difference δ given by

$$\delta = \frac{2\pi}{\lambda}(s_2P - s_1P) = \frac{2\pi}{\lambda}\Delta, \quad (4.1-1)$$

where λ represents the wavelength. It is assumed that the waves start out from s_1 and s_2 in the same phase, because these slits were taken to be equidistant from the source slit s . Furthermore, the magnitudes are practically the same if s_1 and s_2 are of equal width and very close together.

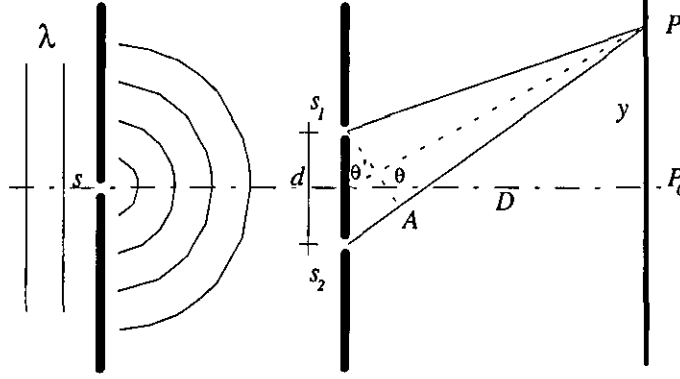


Fig. 4.3 - Path difference in Young's experiment.

The problem of finding the resultant intensity at P therefore reduces to that discussed in Section 1.4, where it was considered the sum of two optical fields. So, the two waves have the same frequency and intensity I_0 , but of phase difference δ . The intensity in P is given by Eq. (1.4-4) as

$$I = 2I_0 + 2I_0 \cos \delta = 2I_0(1 + \cos \delta) = 4I_0 \cos^2\left(\frac{\delta}{2}\right). \quad (4.1-2)$$

It now remains to evaluate the phase difference in terms of the distance y on the screen from the central point P_0 , the separation d between the two slits, and the distance D from the slits to the screen. The corresponding path difference is the distance s_2A in Fig. 4.3, where the dashed line s_1A has been drawn to make s_1 and A equidistant from P . As Young's experiment is usually performed, D is some thousand times larger than d or y . Hence the angles θ and θ' are very small and practically equal. Under these conditions, s_1As_2 may be regarded as a right triangle, and the path difference becomes $d \sin \theta' \approx d \sin \theta$. To the same approximation, one may set the sine of the angle equal to the tangent, so that $\sin \theta \approx y/D$. With these assumptions, one can obtain

$$\Delta = d \sin \theta = d \frac{y}{D}. \quad (4.1-3)$$

This is the value of the path difference to be substituted in Eq. (4.1-1) to obtain the phase difference δ . Now Eq. (4.1-2) for the intensity has maximum values equal to $4I_0$.

whenever δ is an integer multiple of 2π , and according to Eq. (4.1-1) this will occur when the path difference is an integer multiple of λ . Hence one can have

$$d \frac{y}{D} = 0, \lambda, 2\lambda, \dots = m\lambda \quad (4.1-4)$$

or

$$y = m\lambda \frac{D}{d} \quad \text{Bright fringes,} \quad (4.1-5)$$

for $m = 0, 1, 2, \dots$. The minimum value of the intensity is zero, and this occurs when $\delta = \pi, 3\pi, 5\pi, \dots$. Then, for these points

$$d \frac{y}{D} = \frac{\lambda}{2}, \frac{3\lambda}{2}, \frac{5\lambda}{2}, \dots = \left(m + \frac{1}{2}\right)\lambda \quad (4.1-6)$$

or

$$y = \left(m + \frac{1}{2}\right)\lambda \frac{D}{d} \quad \text{Dark fringes,} \quad (4.1-7)$$

for $m = 0, 1, 2, \dots$. The whole number m , which characterizes a particular bright fringe, is called the order of interference. According to these equations the distance on the screen between two successive fringes, which is obtained by changing m by unity in either Eq. (4.1-5) or (4.1-7), is constant and equal to $\lambda D/d$. Not only is this equality of spacing verified by measurement of an interference pattern, but one also finds by experiment that its magnitude is directly proportional to the slit-screen distance D , inversely proportional to the separation of the slits d , and directly proportional to the wavelength λ . Knowledge of the spacing of these fringes thus gives a direct determination of the wavelength λ in terms of known quantities.

4.1.3 Fresnel's Biprism

Soon after the double-slit experiment was performed by Young, the objection was raised that the bright fringes he observed were probably due to some complicated modification of the light by the edges of the slits and not true interference. Thus the wave theory of the light was still questioned. Not many years passed, however, before Fresnel brought forward several new experiments in which the interference of two beams of light was

proved in a manner not open to the above objection [1]. One of these, the Fresnel biprism experiment, will be described in some detail.

A schematic diagram of the biprism experiment is shown in Fig. 4.4. The thin double prism P refracts the light from the slit source s into two overlapping beams ac and be . If screens M and N are placed as shown in the figure, interference fringes are observed only in the region bc . A set of closely spaced fringes in the centre of this region is due to interference, while a set of wide fringes at the edges of the pattern is due to diffraction. These wider bands are produced by the vertices of the two prisms, each of which acts as a straightedge, giving the pattern. When the screens M and N are removed from the light path, the two beams will overlap over the whole region ae . For this case the equally spaced interference fringes will be superimposed on the diffraction pattern of a wide aperture.

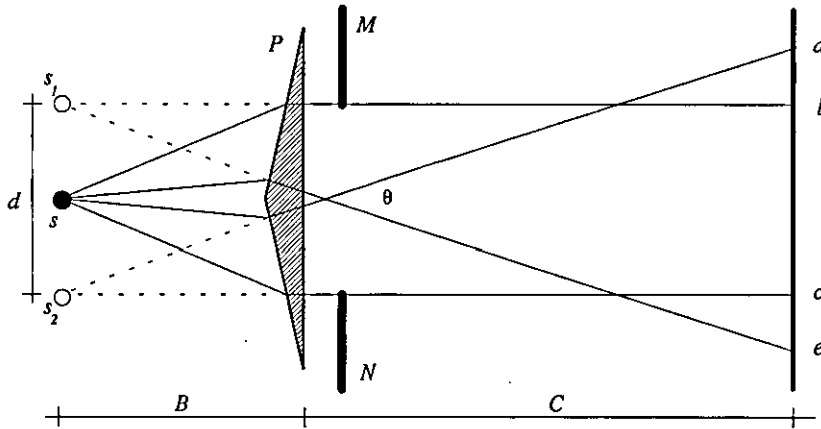


Fig. 4.4 - Diagram of Fresnel's biprism experiment.

With such an experiment Fresnel was able to produce interference without relying upon diffraction to bring the interfering beams together.

Just as in the Young's double-slit experiment, the wavelength of the light can be determined from measurements of the interference fringes produced by the biprism. Calling B and C the distances of the source and screen, respectively, from the prism P , d the distance between the virtual images s_1 and s_2 , and Δy the distance between the successive fringes on the screen, the wavelength λ of the light is given from Eq. (4.1-5) as

$$\lambda = \frac{\Delta y d}{B + C}. \quad (4.1-8)$$

Thus the virtual images s_1 and s_2 act like the two slit sources in Young's experiment.

There is no substantive difference between interference and diffraction. The separation between the two subjects is historical and is retained for pedagogical reasons. Interference may be associated with the intentional formation of two or more light waves that are analyzed in their region of overlap according to the procedures outlined previously. Diffraction may be associated with the obstruction of a single wave, by either a transparent or an opaque obstacle, resulting in the obstruction casting shadows or forming light beams that differ from the size predicted by geometrical optics. The distinction between interference and diffraction is, however, rather fuzzy and arbitrary.

4.2 Fresnel and Fraunhofer Diffraction

When a beam of light passes through a narrow slit, it spreads out to a certain extent into the region of the geometrical shadow. This effect, already mentioned in the beginning of this chapter, is one of the simplest examples of diffraction, that is, of the failure of light to travel in straight lines. It can be satisfactorily explained only by assuming a wave character for light, and this section investigates quantitatively the diffraction pattern, or distribution of intensity of the light behind the aperture, using the principles of wave motion.

Diffraction phenomena are conveniently divided into two general classes [4]:

- those in which the source of light and the screen on which the pattern is observed are effectively at infinite distances from the aperture causing the diffraction;
- those in which either the source or the screen, or both, are at finite distances from the aperture.

The phenomena coming under the first class are called, for historical reasons, Fraunhofer diffraction, and those coming under the second class are called Fresnel diffraction. Fraunhofer diffraction is much simpler to treat theoretically. It is easily observed in practice by rendering the light from a source parallel with a lens and focusing it on a screen with another lens placed behind the aperture, an arrangement which effectively removes the source and screen to infinity. In the observation of Fresnel diffraction, on the other hand, no lenses are necessary, but here the wave-fronts are divergent instead of plane, and the theoretical treatment is consequently more complex.

Only Fraunhofer diffraction will be considered throughout this thesis because diffraction pattern calculations can be reduced to simple mathematical manipulations. In particular,

the observed field distribution of a diffraction pattern can be found directly from a Fourier transform of the aperture distribution itself [5]. The application program, which has been developed and described in chapter two, can be used to calculate and generate Fraunhofer diffraction patterns on the computer screen for different aperture distributions.

In the next sections a few examples of Fraunhofer diffraction will be presented and described before the discussion about the image formation process. The results can be applied directly to find the complex distribution across the Fraunhofer diffraction pattern of any given aperture. However, it should be noted that real radiation detectors respond only to optical intensity rather than field magnitude. The final description of the diffraction patterns will be a distribution of intensity. A reference to the importance of the phase associated with the diffraction pattern in the image reconstruction process will also be made.

4.2.1 Diffraction by a Single Slit

A slit is a rectangular aperture of length large compared to its width. Consider a slit s to be set up as shown in Fig. 4.5, with its long dimension perpendicular to the plane yz , and to be illuminated by parallel monochromatic light from the narrow slit s' , at the principal focus of the collimating lens L_1 . The light focused by the objective lens L_2 on a screen or photographic film P at its principal focus will form a diffraction pattern. Since the Fraunhofer diffraction pattern of a single slit, s , is the Fourier transform of its aperture the focal plane of the objective lens L_2 is normally called the Fourier or diffraction plane of the optical system.

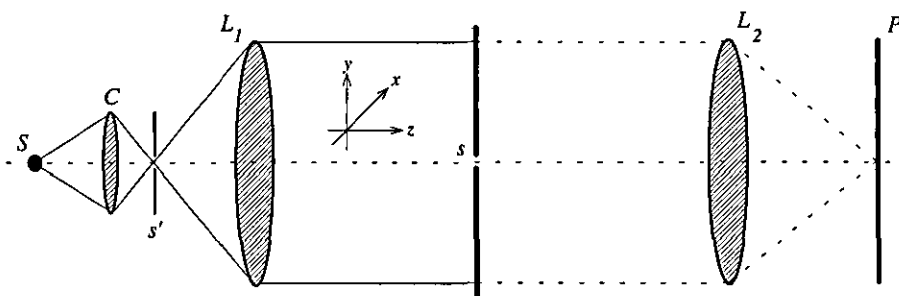


Fig. 4.5 - Experimental arrangement for obtaining the diffraction pattern of a single slit; Fraunhofer diffraction.

Figure 4.6(a) shows the image of the considered aperture with a constant distribution of intensity. In Fig. 4.6(b) and (c) is shown the resulting diffraction pattern intensity and

one vertical profile, respectively. It is important to note that the width of the central maximum is twice as great as that of the fainter side maxima.

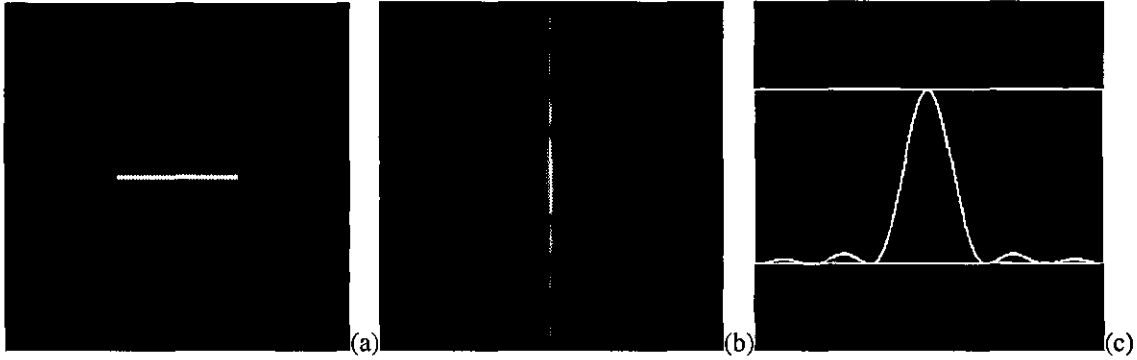


Fig. 4.6 - The single slit diffraction pattern. (a)- slit aperture; (b)- diffraction pattern; (c)- vertical profile of (b).

The explanation of the single slit pattern lies in the interference of the Huygens secondary wavelets which can be thought of as sent out from every point on the wave-front at the instant that it occupies the plane of the slit. To a first approximation, one may consider these wavelets to be uniform spherical waves, the emission of which stops abruptly at the edges of the slit. The results obtained in this way, although they give a fairly accurate account of the observed facts, are subject to certain modifications in the light of the more rigorous theory.

The displacement in the y direction of a wave travelling in the z direction can be expressed by the following equation

$$y = m \sin(\omega t - kz), \quad (4.2-1)$$

where m is the magnitude, $\omega = 2\pi\nu$ is the angular frequency of the oscillation, and $k = 2\pi/\lambda$ is the wave number. Figure 4.7 represents a section of a slit of width b , illuminated by a parallel beam of light from the left. Let ds be the width of an element of the wave-front in the plane of the slit, at a distance s from the origin O . The parts of each secondary wave which travel normal to the plane of the slit will be focused at P_0 , while those which travel at any angle θ will reach P . Considering first the wavelet emitted by the element ds situated at the origin, its magnitude will be directly proportional to the length ds and inversely proportional to the distance z . At P it will produce an infinitesimal displacement which, for a spherical wave, may be expressed as

$$dy_0 = \frac{m ds}{z} \sin(\omega t - kz). \quad (4.2-2)$$

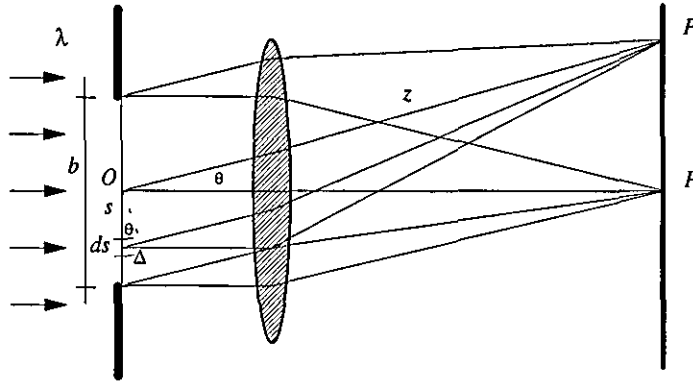


Fig. 4.7 - Geometrical construction for investigating the intensity in the single slit diffraction pattern.

As the position of ds is varied, the displacement it produces will vary in phase because of the different path length to P . When it is at a distance s below the origin, the contribution will be

$$dy_{-s} = \frac{m ds}{z} \sin[\omega t - k(z + \Delta)] = \frac{m ds}{z} \sin(\omega t - kz - ks \sin \theta). \quad (4.2-3)$$

The effects of all the elements from one edge of the slit to the other must be added together. This can be done by integrating Eq. (4.2-3) from $s = -b/2$ to $b/2$. The simplest way is to integrate the contributions from pairs of elements symmetrically placed at $-s$ and s , each contribution being

$$\begin{aligned} dy &= dy_{-s} - dy_s \\ &= \frac{m ds}{z} [\sin(\omega t - kz - ks \sin \theta) + \sin(\omega t - kz + ks \sin \theta)] \end{aligned} \quad (4.2-4)$$

By the identity $\sin \alpha + \sin \beta = 2 \cos \frac{1}{2}(\alpha - \beta) \sin \frac{1}{2}(\alpha + \beta)$, one can have

$$dy = \frac{m ds}{z} [2 \cos(ks \sin \theta) \sin(\omega t - kz)], \quad (4.2-5)$$

which must be integrated from $s = 0$ to $b/2$. In doing so, z may be regarded as constant.

Thus

$$\begin{aligned} y &= \frac{2m}{z} \sin(\omega t - kz) \int_0^{b/2} \cos(ks \sin \theta) ds \\ &= \frac{2m}{z} \left[\frac{\sin(ks \sin \theta)}{k \sin \theta} \right]_0^{b/2} \sin(\omega t - kz) \\ &= \frac{mb \sin(\frac{1}{2} kb \sin \theta)}{z \frac{1}{2} kb \sin \theta} \sin(\omega t - kz) \end{aligned} \quad (4.2-6)$$

The resultant vibration will therefore be a simple harmonic one, the magnitude of which varies with the position of P , since the latter is determined by θ . The magnitude can be represented by the following expression

$$M = M_o \frac{\sin \beta}{\beta}, \quad (4.2-7)$$

where $\beta = \frac{1}{2}kbs \sin \theta = (\pi bs \sin \theta)/\lambda$ and $M_o = mb/z$. The quantity β is a convenient variable, which signifies one-half the phase difference between the contributions coming from opposite edges of the slit. The intensity on the screen, which is by definition the square of the magnitude, is then

$$I \approx M^2 = I_o \frac{\sin^2 \beta}{\beta^2}, \quad (4.2-8)$$

where $I_o = M_o^2$. The intensity curve will be seen to have the form required by the experiment result shown in Fig. 4.6(b) and (c). The maximum intensity of the strong central band comes at the point P_o of Fig. 4.7, where evidently all the secondary wavelets will arrive in phase because the path difference is $\Delta = 0$. For this point $\beta = 0$, and although the quotient $(\sin \beta)/\beta$ becomes indeterminate for $\beta = 0$, it will be remembered that $\sin \beta$ approaches β for small angles and is equal to it when β vanishes. Hence for $\beta = 0$, $(\sin \beta)/\beta = 1$. One can see the significance of the constant M_o . Since for $\beta = 0$, $M = M_o$, it represents the magnitude when all the wavelets arrive in phase. M_o^2 is then the value of the maximum intensity, at the centre of the pattern. From this principal maximum the intensity falls to zero at $\beta = \pm\pi$, then passes through several secondary maxima, with equally spaced points of zero at $\beta = \pm\pi, \pm2\pi, \pm3\pi, \dots$. The secondary maxima do not fall halfway between these points, but are displaced toward the centre of the pattern by an amount which decreases with increasing β .

A very clear idea of the origin of the single slit pattern is obtained by the following simple treatment. Consider the light from the slit, shown in Fig. 4.8, coming to the point P_1 on the screen, this point being just one wavelength farther from the upper edge of the slit than from the lower. The secondary wavelet from the point in the slit adjacent to the upper edge will travel approximately $\lambda/2$ farther than that from the point at the centre, and so these two will produce vibrations with phase difference of π and will give a resultant displacement of zero at P_1 . Similarly the wavelet from the next point below the upper edge will cancel that from the next point below the centre, and one can continue this pairing off to include all points in the wave-front, so that the resultant effect at P_1 is zero.

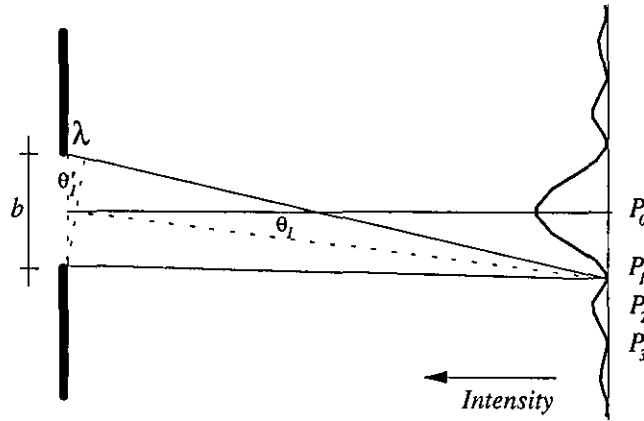


Fig. 4.8 - Angle of the first minimum of the single slit diffraction pattern.

At P_3 the path difference is 2λ , and if one divide the slit into four parts, the pairing of points again gives zero resultant, since the parts cancel in pairs. For the point P_2 , on the other hand, the path difference is $3\lambda/2$, and dividing the slit into thirds, two of which will cancel, leaving one third to account for the intensity at this point. The resultant magnitude at P_2 is, of course, not even approximately one-third that at P_0 , because the phases of the wavelets from the remaining third are not by any means equal.

The above method, though instructive, is not exact if the screen is at a finite distance from the slit. As Fig. 4.8 is drawn, the shorter broken line is drawn to cut off equal distances on the rays to P_1 . It will be seen from this that the path difference to P_1 between the light coming from the upper edge and that from the centre is slightly greater than $\lambda/2$ and that between the centre and the lower edge slightly less than $\lambda/2$. Hence the resultant intensity will not be zero at P_1 and P_3 , but it will be more nearly so the greater the distance between slit and screen or the narrower the slit. This corresponds to the transition from Fresnel diffraction to Fraunhofer diffraction. Obviously, with the relative dimensions shown in the figure, the geometrical shadow of the slit would considerably widen the central maximum as drawn. Just as was true with Young's experiment, when the screen is at infinity, the relations become simpler. Then the two angles θ_1 and θ'_1 in Fig. 4.8 become exactly equal, that is, the two broken lines are perpendicular to each other, and $\lambda = b \sin \theta_1$ for the first minimum corresponding to $\beta = \pi$. This gives

$$\sin \theta_1 = \frac{\lambda}{b}. \quad (4.2-9)$$

In practice θ_1 is usually a very small angle, so one may put the sine equal to the angle. Then $\theta_1 = \lambda/b$, is a relation that shows at once how the dimensions of the pattern vary with λ and b . The linear width of the pattern on the screen will be proportional to the

slit-screen distance, which is the focal length f of a lens placed close to the slit. The linear distance d between successive minima corresponding to the angular separation $\theta_1 = \lambda/b$ is thus

$$d = \frac{\lambda f}{b}. \quad (4.2-10)$$

The width of the pattern increases in proportion to the wavelength, so that for red light it is roughly twice as wide as for violet light, the slit width, etc., being the same. The angular width of the pattern for a given wavelength is inversely proportional to the slit width b , so that as b is made larger, the pattern shrinks rapidly to a smaller scale. Then, when the width of the aperture is large compared with the wavelength, the diffraction effect is practically negligible.

4.2.2 Diffraction by a Rectangular Aperture

In the preceding section the intensity function for a slit was derived by summing the effects of the spherical wavelets originating from a linear section of the wave-front by a plane perpendicular to the length of the slit, that is, by the plane yz in Fig. 4.5. Nothing was said about the contributions from parts of the wave-front out of this plane. A more detailed mathematical investigation, involving a double integration over both dimensions of the wave-front [6], shows, however, that the above result is correct when the slit is very long compared to its width.

In this section the intensity function of the diffraction pattern of a rectangular aperture will be derived directly using Fourier transformations. Consider first a rectangular aperture of sides $2a$ and $2b$ with origin o at the centre of the rectangle and with ox and oy axes parallel to the sides as shown in Fig. 4.9. The light distribution over the aperture is assumed constant and equal to M_o .

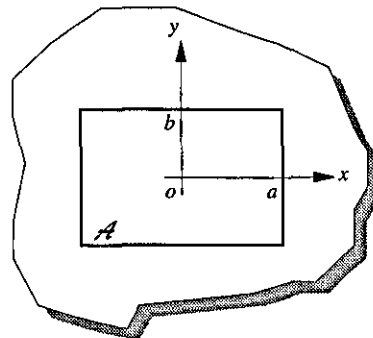


Fig. 4.9 - Rectangular aperture.

The integral that governs the Fraunhofer diffraction pattern [7] is a complex valued function $F(u, v)$ defined by

$$F(u, v) = \iint_{\mathcal{A}} f(x, y) \exp[-jk(xu + yv)] dx dy, \quad (4.2-11)$$

where $f(x, y)$ is an aperture distribution function, (u, v) is the co-ordinate pair on the diffraction screen, and $k = 2\pi/\lambda$ is the wave number. Comparing this equation with Eq. (3.2-9) in chapter three, one can observe that Eq. (4.2-11) expresses the Fraunhofer diffraction pattern function $F(u, v)$ as the Fourier transform of the aperture distribution $f(x, y)$. For this case in particular, one can rewrite the integral as

$$\begin{aligned} F(u, v) &= M_o \int_{-a}^a \int_{-b}^b \exp[-jk(xu + yv)] dx dy \\ &= M_o \int_{-a}^a \exp[-jkxu] dx \int_{-b}^b \exp[-jkyv] dy. \end{aligned} \quad (4.2-12)$$

Now

$$\int_{-a}^a \exp[-jkxu] dx = 2a \frac{\sin(kau)}{kau}, \quad (4.2-13)$$

with a similar expression for the other integral. Hence the intensity is given by

$$I = |F|^2 = I_o \frac{\sin^2 \beta}{\beta^2} \frac{\sin^2 \gamma}{\gamma^2}, \quad (4.2-14)$$

where $I_o = (4ab)^2 M_o^2$ is the intensity at the centre of the pattern, $\beta = kau$ and $\gamma = kbv$.

From Eq. (4.2-14) one can see that the intensity is zero along two sets of lines parallel to the sides of the rectangle, given by $\beta, \gamma = \pm\pi, \pm2\pi, \pm3\pi, \dots$. Within each rectangle formed by pairs of consecutive dark lines the intensity rises to a maximum. All these maxima are, however, only a small fraction of the central maximum, and decrease rapidly with increasing distance from the centre, as shown in Fig. 4.10. The larger the opening, the smaller is the effective size of the diffraction pattern.

Figure 4.10 shows, as an example, the intensity distribution of the Fraunhofer diffraction pattern function of a rectangular aperture which have been calculated using its Fourier transformation. In particular, Fig. 4.10(a) shows the intensity distribution function of the light through the aperture and in (b) the corresponding diffraction

pattern. Figure 4.10(c) and (d) shows one vertical and one horizontal profile of the intensity distribution shown in (b), respectively.

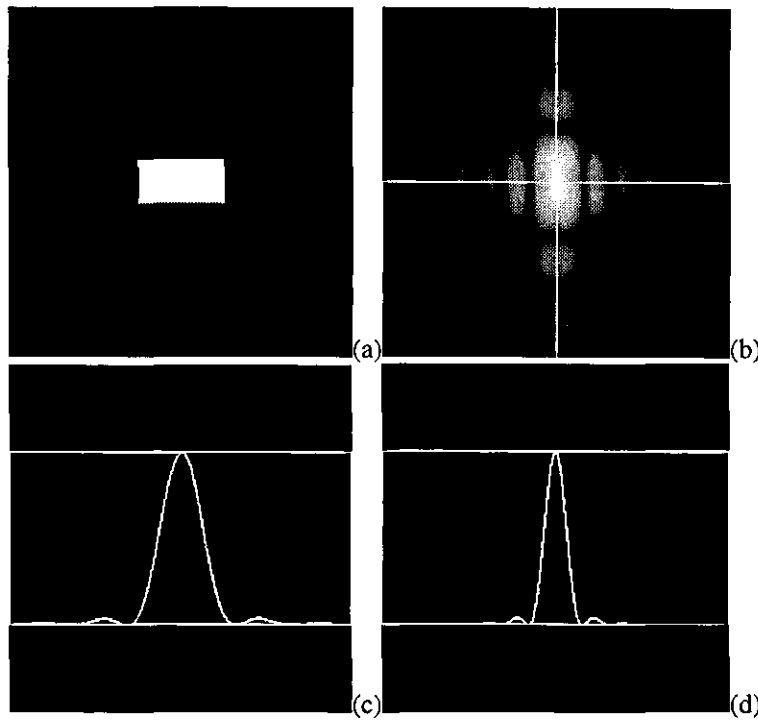


Fig. 4.10 - Fraunhofer diffraction pattern of a rectangular aperture. (a)- rectangle aperture; (b)- diffraction pattern; (c)- vertical profile marked in (b); (d)- horizontal profile marked (b).

The intensity in the diffraction pattern is concentrated principally in two directions coinciding with the sides of the aperture, and in each of these directions it corresponds to the simple pattern for a slit width equal to the width of the aperture in that direction. Owing to the inverse proportionality between the slit width and the scale of the pattern, the fringes are more closely spaced in the direction of the longer dimension of the aperture. In addition to these patterns there are other weak maxima in the field, as shown in the above figure.

4.2.3 Diffraction by a Circular Aperture

In a similar way one can investigate the Fraunhofer diffraction pattern at a circular aperture. Consider first a circular aperture of radius a with origin o at the centre of the circle and with ox and oy axes perpendicular to the circle as shown in Fig. 4.11. The light distribution through the aperture is assumed constant and equal to M_o .

It is now appropriate to use polar instead of rectangular co-ordinates. Let (ρ, θ) be the polar co-ordinate pair of a typical point in the aperture.

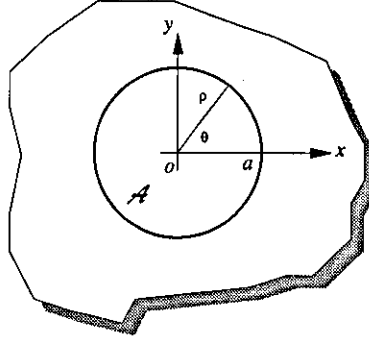


Fig. 4.11 - Circular aperture.

These polar co-ordinates are related to the rectangular co-ordinates x and y by the following relations

$$\begin{aligned} x &= \rho \cos \theta \\ y &= \rho \sin \theta \end{aligned} \quad (4.2-15)$$

Let also (η, ξ) be the polar co-ordinates of a point in the diffraction pattern, referred to the geometrical image of the source by

$$\begin{aligned} u &= \eta \cos \xi \\ v &= \eta \sin \xi \end{aligned} \quad (4.2-16)$$

From the definition of u and v , it follows that $\eta = \sqrt{u^2 + v^2}$ is the sine of the angle which the direction (u, v) makes with the central direction $u = v = 0$. The diffraction integral, as defined in Eq. (4.2-11), after the variable change becomes

$$F(\eta, \xi) = M_o \int_0^a \int_0^{2\pi} \exp[-jk\rho\eta \cos(\theta - \xi)] \rho d\rho d\theta. \quad (4.2-17)$$

Now, one has the well-known integral representation of the Bessel functions $J_n(z)$ [8]:

$$J_n(z) = \frac{j^{-n}}{2\pi} \int_0^{2\pi} \exp[jz \cos \alpha] \exp[jn\alpha] d\alpha. \quad (4.2-18)$$

Equation (4.2-17) therefore, for $n = 0$, reduces to

$$F(\eta, \xi) = 2\pi M_o \int_0^a J_0(k\rho\eta) \rho d\rho. \quad (4.2-19)$$

Also, there is the well-known recurrence relation [9]:

$$\frac{d}{dz}[z^{n+1}J_{n+1}(z)] = z^{n+1}J_n(z), \quad (4.2-20)$$

given, for $n = 0$, on integration the following relation

$$zJ_1(z) = \int_0^z sJ_0(s)ds. \quad (4.2-21)$$

From Eq. (4.2-19) one can make a variable change $s = k\rho\eta$, so $\rho = s/(k\eta)$ and $d\rho = ds/(k\eta)$ given the new representation for the integral as

$$F(\eta, \xi) = \frac{2\pi M_o}{(k\eta)^2} \int_0^{k\eta} sJ_0(s)ds. \quad (4.2-22)$$

Using the relation expressed in Eq. (4.2-21) the diffraction integral in Eq. (4.2-22) becomes

$$F(\eta, \xi) = \pi a^2 M_o \left[\frac{2J_1(k\eta)}{(k\eta)} \right]. \quad (4.2-23)$$

Hence the intensity is given by

$$I = |F|^2 = I_o \left[\frac{2J_1(k\eta)}{(k\eta)} \right]^2, \quad (4.2-24)$$

where $I_o = (\pi a^2)^2 M_o^2$. This is a celebrated formula first derived in a somewhat different form by the astronomer Sir George Airy in 1835 [10].

The intensity distribution in the neighbourhood of the geometrical image is characterized by the function defined in Eq. (4.2-24) and shown in Fig. 4.12. In particular, Fig. 4.12(a) shows the intensity distribution function of the light through the aperture and in (b) the corresponding Airy diffraction pattern. Also, Fig. 4.12(c) shows one profile of the diffraction pattern. It can be seen that it has its principal maximum $I = I_o$ at $\eta = 0$, and with the increasing η it oscillates with gradually diminishing amplitude, in a similar way to the function $(\sin\beta/\beta)^2$ discussed in Section 4.2.2. The intensity is zero (minimum) for values of η given by $J_1(k\eta) = 0$. The minima are no longer strictly equidistant and with increasing η the separation between two successive minima or two successive maxima approaches the value π .

The results displayed in Fig. 4.12 shows that the pattern consists of a bright disc, centred on the geometrical image $u = v = 0$ of the source, surrounded by concentric bright and dark rings.

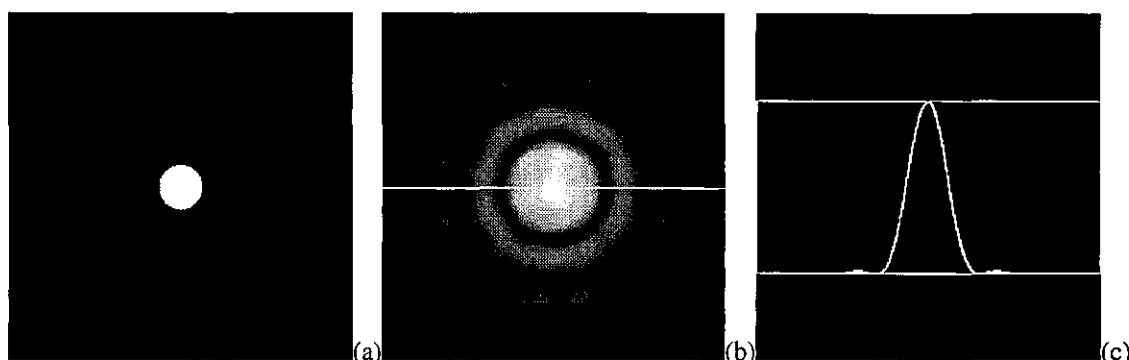


Fig. 4.12 - Fraunhofer diffraction pattern of a circular aperture (the Airy pattern). (a)- circular aperture; (b)- diffraction pattern; (c)- profile marked in (b).

The intensity of the bright rings decreases rapidly with their radius and normally only the first one or two rings being bright enough to be visible to the naked eye. Using a table of the Bessel functions it follows that the radii of the dark rings are

$$\eta = \sqrt{u^2 + v^2} = 0.610 \frac{\lambda}{a}, \quad 1.116 \frac{\lambda}{a}, \quad 1.619 \frac{\lambda}{a}, \dots \quad (4.2-25)$$

The separation between two neighbouring rings approaches asymptotically the value $\lambda/(2a)$. The effective size of the diffraction pattern is again seen to be inversely proportional to the linear dimensions of the aperture.

Fraunhofer diffraction at apertures of other forms may be studied in a similar manner, the calculations being particularly simple when the aperture distributions functions are discrete functions that are suitable for storage in a digital computer where fast Fourier transforms can be efficiently calculated, as mentioned in chapter three, using appropriate algorithms.

4.3 Image Formation

The major discoveries attributed to the wave properties of light, diffraction, interference and polarization, were all accounted for over 100 years ago. Up until the beginning of the 20th century, nearly all optical phenomena had been studied by Fresnel, Fraunhofer, Huygens, Airy, Foucault, Young, and a few others. Over the years these fundamental

principles have found many practical applications in the development of microscopes, binoculars, periscopes, telescopes, interferometers, etc. Detailed studies of diffraction phenomena have led in recent years to the development of a large number of useful optical instruments which can be best described using the wave picture of light.

In the image formation process radiation is used as a vehicle for conveying information about an object to the place where the radiation is distributed in such a way as to form an image of the object. With a suitable lens the light leaving the two slits in Young's experiment, in Section 4.1.2, can be used to produce just such an image of the slits. Similarly, when the slits are viewed directly by the eye it is the lens in the eye that produces an image on the retina at the back of the eye. How does the lens function? Why is it that with a large source at s , see Fig. 4.3, fringes are not seen on the screen yet the insertion of the same lens will still produce an image of them?

The answers to these questions involve not only diffraction and interference but also the coherence of the radiation [11]. For interference effects to be observed on the screen in Young's experiment it is necessary that the wave-trains of the light arriving there from the two slits should overlap and of course have the same frequency, and that there should be a constant phase difference between them. If these conditions were ideally met, the illumination at the apertures would be said to be coherent. Conventional light sources are not perfect, however, they provide illumination that is to a greater or lesser extent coherent.

Consider the first requirement for coherence mentioned above. From each wave-front incident at the aperture containing the two slits in Young's experiment, new wave-fronts spread out from s_1 and s_2 . One visualizes the parts of the wave-fronts that travel the paths from s_1 and s_2 to P as wave-trains, see Fig. 4.3. If the wave-trains are to interfere at P they must not only be of the same frequency but must overlap there, and this will only happen if their length is large compared with the difference in the distance they travel, otherwise they could miss meeting each other at P . One can say that the light must have sufficient coherence length, $l_c = ct_c$, where t_c is called the coherence time of the illumination and c the velocity of the light. This is a description of the temporal coherence of the illumination and it is related to the spectral composition of the light [12]. Coherence time and coherence length of the illumination used in interference experiments is a key factor in determining fringe visibility.

For the second requirement, that there be a constant phase difference between the wave-trains arriving at s_1 and those arriving at s_2 , the light source must be small enough to

ensure that there is a negligible spread, compared with the wave-length of the light, of the distances travelled from different points in the source, to s_1 and s_2 . This is the requirement for spatial coherence of the illumination field over the width $d = \overline{s_1 s_2}$ [12].

It was noted that in Young's experiment the introduction of a lens would enable an image of the two slits to be formed. It is evident, then, that diffraction at the slits is the first step in the formation of their optical image, the second step being the recombination of the light by the lens to form the image.

Does the lens act differently when a large source of light is used? The short answer is no, for reasons that can fairly readily be appreciated in qualitative terms.

Consider the first case, when the Young's experiment is performed in the classical way with what one can describe as reasonably coherent illumination at the aperture screen. The source is small enough to ensure that the same phase relationships exist between every pair of diffracted wave-fronts leaving the slits. every pair contributes to the same set of fringes, and the intensity maxima and minima in the fringe pattern can be located by reference to Eq. (4.1-2).

In the second case, using a large source, there is no spatial coherence in the illumination at the aperture screen. The fringes resulting from wave-fronts leaving one point in the source are displaced with respect to fringes produced by wave-fronts originating from other points in the source. If the source is large enough the net result is a relatively uniform patch of illumination on the screen. Even so, fringes are indeed formed as in the first case, but they are instantaneous and their positions continuously change, given uniform illumination when viewed over even the shortest practicable time. That the information is there, with individual phase relationships still existing in the light that will enter the lens when this is inserted, is evident from the fact that it is still possible to form an image of the slits with the lens.

The above comments concerning the imaging under coherent illumination, usually called coherent imaging, of an object in the form of the slit in Young's experiment, apply equally to more complicated objects such as the 35-mm transparencies used in slide projectors, opaque objects illuminated with conventional thermal light, and self-luminous objects which are luminescent (TV picture) or incandescent (infra-red photography). The same instantaneous phase relationships exist in each of these categories, just as described above. In this wider context of imaging the term scattering

is often used instead of diffraction. There is some confusion in making a distinction between the two in the literature and one can regard the terms as synonymous.

There are two aspects of image formation that now need to be noted. One is concerned with the formation of a Fraunhofer-type diffraction pattern in the back focal plane of the imaging lens, the other with the effect that a finite aperture of the lens has on the image. They lead to two useful models of image formation as described below.

Consider first Fig. 4.13 which depicts an object mask containing two very small slit apertures, s_1 and s_2 , uniformly illuminated with quasi-monochromatic light from a distant source. Plane waves arrive at normal incidence to the object mask, and spherical wave-fronts spread from s_1 and s_2 .

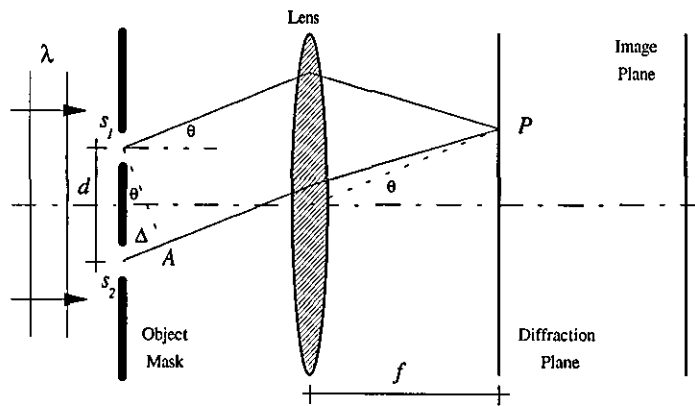


Fig. 4.13 - Diffraction as an intermediate step in image formation.

This is the same as in Young's experiment except that in addition it has a lens producing an image of the slits in a plane located as shown in the figure. It is the back focal plane of the lens, however, that is of immediate interest. Consider any point P , at θ to the lens axis. Only contributions travelling in direction θ from s_1 and s_2 are brought together there, to interfere. The particular diffraction pattern, defined below as of the Fraunhofer-type, formed in the back focal plane of the imaging lens, is a particularly important intermediate step in the formation of the image produced by the lens. It provides a valuable insight into the final stage of image formation, and an opportunity of special and unique significance for modifying the image.

It should be noted, however, that for such Fraunhofer diffraction patterns to be experimentally accessible it is necessary to arrange that the static phase relationships provided by coherent illumination exist, as it was mentioned before. To calculate the resultant of the two contributions at P in Fig. 4.13, it is necessary to take into account the phase difference, δ , between them, as described in Section 4.1.2, due to the path

difference Δ . Regardless of the shape, number, etc., of apertures in the object mask, the patterns formed in this way are all described as being of the Fraunhofer-type.

Several arrangements can be used specifically for observing and recording Fraunhofer diffraction patterns [5,11,12], often for optical filtering purposes, but when accurate details are required concerning both phase and magnitude, the requirements are more stringent. This is due to the imaging properties of the lenses, which when fully taken into account show that the arrangement shown in Fig. 4.14, where the object is located in the front focal plane of the lens in whose back focal plane the diffraction pattern is formed, should be used [5,12].

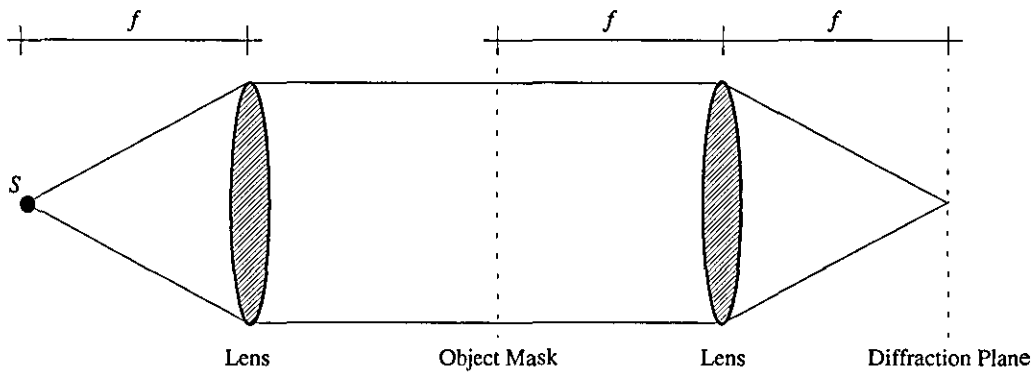


Fig. 4.14 - Arrangement for forming Fraunhofer diffraction patterns.

It should also be noted (Fig. 4.14) that the Fraunhofer diffraction pattern given by any object is observed in the plane where the image of the light source is formed. Using the term conjugate in the same sense as in geometrical optics, where one refers to an image as being formed in the plane conjugate to the object, the Fraunhofer diffraction pattern is formed in the plane conjugate with the light source, as clearly shown in the models of image formation presented in Fig. 4.15.

Since the wave-front system leaving an object is the only information on which the formation of its image is based it is to be expected, and is indeed found, that the more of the wave-front system that is allowed to enter the imaging lens the better is the quality of the image. Rephrased, this is the well-known axiom that the larger the aperture of the lens the better is the definition of the image.

The first model, identified by a continuous line (i) in Fig. 4.15, places emphasis on the overall diffraction (scattering) of the light from an object when the conditions are at least partially coherent, and the way in which that light is recombined to form an image. The interpretation of the Fraunhofer diffraction pattern given by any object function as the Fourier transform of that object function is a most valuable concept and it should be

used freely. It is interesting to realize that it leads to the description of a lens as a Fourier transformer and for that reason the plane of a diffraction pattern is normally referred to as the Fourier plane (or Fourier domain), or alternatively as the spatial frequency plane (or spectral domain).

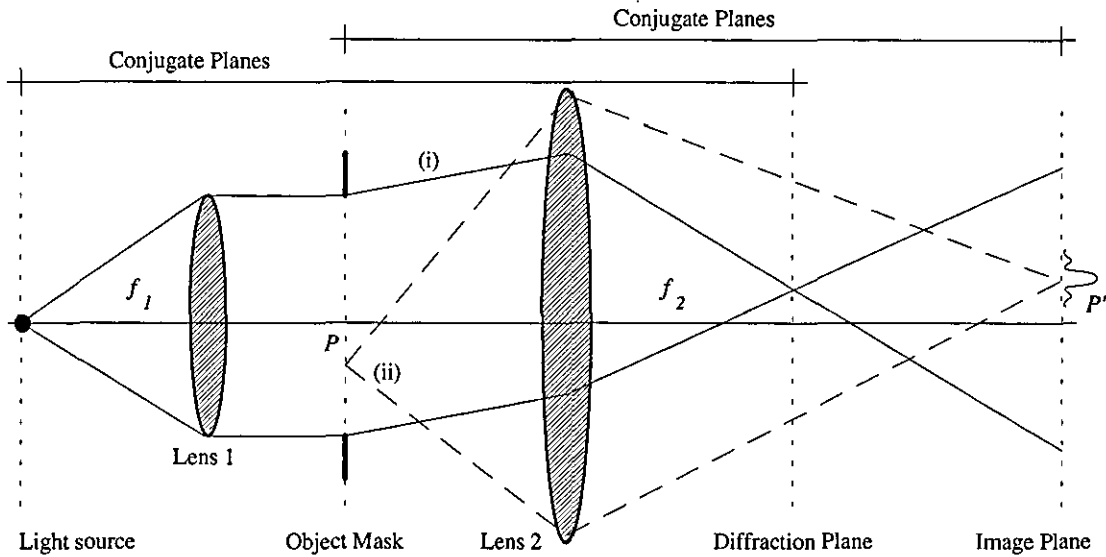


Fig. 4.15 - Two models of image formation.

By contrast, the plane where an image is formed is normally referred to as the image plane (or image domain), or alternatively as the space plane (or spatial domain) which is the conjugate plane of the object. This model was originally formulated essentially qualitatively by Abbe in 1873 [13] to deal with his experimental observations on the imaging of periodic objects under the microscope. For a coherently illuminated object the quantitative identification of diffraction orders with the terms of the Fourier series analysis describing its structure, and the way an imaging lens recombines them as a Fourier synthesis to form an image, were reported in some detail by Porter in 1906 [14]. The extension to the Abbe-Porter interpretation to include the analogous role of Fourier transforms in the coherent imaging of non-periodic objects was made mainly by X-ray crystallographers by the early 1940s. The value of the approach offered by this model is its sensitivity to the way in which spatial frequencies in an object structure (periodic or non-periodic) are conveyed by diffracted wave-fronts and reconstructed to form an image. Using coherent illumination it leads to methods for intervention in the diffraction plane (the spatial frequency plane) so that the formation of the image can be controlled by spatial filtering which is one important aspect in optical processing.

The second model of image formation, identified by a dashed line (ii) in Fig. 4.15, is applicable to both coherent and incoherent illumination conditions. Here Rayleigh in 1896 [15] made an important contribution with influences from the early work of Airy

[10]. The model visualizes an image as the combination of the Airy patterns, or more complicated patterns, that an optical system would separately produce with the light leaving every individual point in the object. As noted before, the Fraunhofer diffraction patterns given by aperture masks are formed in the plane where the image of the light source is formed, that is, the plane conjugate to the light source. In the present instance, in the model of imaging developed by Rayleigh, each point in an object being imaged is itself considered as a point source of light. The wave-fronts from it are limited in their entry into the imaging system by the finite aperture of the imaging lens, and the diffraction pattern of that aperture is formed in the plane conjugate to the plane containing the point source, that is, in the image plane of the object points as light sources (Fig. 4.15). Each point of the object is therefore imaged not as a point but as the Airy pattern of the aperture of the imaging lens, or a more complicated pattern. If the illumination is incoherent the Airy intensity patterns due to all the object points are simply additive. If it is coherent there is interference and mathematically one deals with the combination of the complex amplitude Airy patterns. An important development of this model of image formation is due to Duffieux [16] who appears to have been the first to formally express the image of a continuous object distribution as a convolution of that distribution with the impulse response of the system, which is normally called the point-spread function, using complex amplitudes or intensities as appropriate for the illumination conditions. As mentioned in chapter three, application of the convolution theorem then showed that the spatial frequency spectrum of the image is the product of the spatial frequency spectrum of the object distribution and that of the response of the system.

In the remainder of this thesis it can be assumed, unless otherwise stated, that the various diffraction patterns discussed are of the Fraunhofer-type or that they approximate sufficiently to it for the differences to be negligible for the purposes concerned. Also, the first model of image formation previously described will be used for all the simulations and experimental results presented in the following chapters.

4.4 Image Reconstruction

Suppose one knows that two functions, $f(x,y)$ and $F(u,v)$, form a Fourier transform pair, meaning that $F(u,v)$ is the Fourier transform of $f(x,y)$, and $f(x,y)$ is the inverse Fourier transform of $F(u,v)$, where the relation can be identified by

$$f(x,y) \Leftrightarrow F(u,v). \quad (4.4-1)$$

It is now appropriate to call $f(x,y)$ the image and $F(u,v)$ its diffraction pattern, with (x,y) and (u,v) being the rectangular co-ordinates of arbitrary points in the two-dimensional image space (image domain) and in the two-dimensional Fourier space (Fourier domain). Both the image and the diffraction pattern are in general complex valued functions, and the notation

$$Z = |Z| \exp[j\phi] \quad (4.4-2)$$

is a convenient way for identifying the magnitude $|Z|$ and the phase ϕ of any complex quantity Z , whose complex conjugate and intensity are written as Z^* and $|Z|^2$ respectively.

Suppose further that one is presented with merely the diffraction pattern magnitude $|F(u,v)|$. What can be inferred about its phase, say $\psi(u,v)$, in the absence of extra information? Can one come to sharper conclusions when the image is subject to particular constraints? Such questions are aspects of what is called the Fourier phase problem.

Mathematically educated people who come upon this problem for the first time tend to be surprised that it can have much content. After all, why should one not attach *any old* phase to the given magnitude? Such impressions are strongly reinforced by the well-known *dominance* of phase over the magnitude, as illustrated in the next section. It is nevertheless true that the diffraction pattern phase, also called the spectral phase, is constrained by the diffraction pattern magnitude, or spectral magnitude, in scientifically important applications.

Without any additional information or constraints, spectral magnitude and phase are generally independent functions in the sense that the knowledge of one is not sufficient to uniquely deduce the other. In other words, both magnitude and phase information are generally required in order to uniquely reconstruct the complex valued function. There are a number of important applications in which the ability to reconstruct an image from either the magnitude or the phase of its Fourier transform would be useful. The need to recover an image or equivalently to recover spectral phase information from only spectral magnitude information, referred to as the phase retrieval problem, arises in a variety of different contexts and applications such as X-ray crystallography, electron microscopy, coherence theory, optics, and signal processing [17-24]. On the other hand, the ability to recover spectral magnitude from only spectral phase information, referred to as the magnitude retrieval problem, was initially suggested on a theoretical basis in

order to address the importance of phase information [25]. However, in a number of applications the recovery of spectral magnitude would be useful [26].

Due to its theoretical and practical importance, this section addresses the problem of reconstructing a discrete two-dimensional image from either the phase or the magnitude of its Fourier transform. Specifically, two fundamental issues related to phase-only and magnitude-only image reconstruction will be considered.

4.4.1 The Importance of Phase

While a large amount of research effort was being directed toward an understanding of the uniqueness of the solution to the phase retrieval problem and toward the development of techniques for recovering phase information from intensity measurements and a priori signal constraints, it had been noted in a number of different contexts and applications that many of the features of a signal are contained within the phase of its Fourier transform. Specifically, it has been observed that a phase-only synthesis of a signal with a constant or ensemble average magnitude contains a number of similarities to the true signal. It appears that the first context in which the similarity between a signal and its phase-only synthesis was noted is in the field of X-ray crystallography [27]. Since the X-ray diffraction pattern of a crystal is related to the electron density distribution in the crystal by a Fourier transform, if both the magnitude and phase of the diffracted wave could be measured, then an inverse Fourier transform of the diffraction pattern would yield the desired structure. However, since only the intensity of the diffraction pattern may be measured, the only available information which can be obtained from this measurement is the magnitude of the electron density distribution, because it is the square root of the intensity. Therefore, although the magnitude function contains peaks at positions corresponding to the inter atomic vectors between atoms in the crystal, phase information is necessary to uniquely recover the structure. In order to determine the importance of the phase, several experiments in the Fourier synthesis can be performed by combining the correct magnitude with various other phase functions and by combining the correct phase with various other magnitude functions.

The similarity between a signal and its phase-only synthesis has also been noted in the context of image processing by investigators performing a similar set of experiments [25,28]. In particular, it has been observed that when only phase information is used to synthesize an image, the result contains many of the important features of the original

image. An image synthesized on the basis of magnitude information alone, however, will not generally bear any similarity to the original.

Shown in Fig. 4.16, as an example, is a real non-negative original image in which magnitude and phase are displayed in (a) and (b), respectively. Also, the same figure shows respectively in (c) and (d), the corresponding diffraction pattern magnitude and phase, previously calculated using the fast Fourier transform algorithm of the discrete original image.

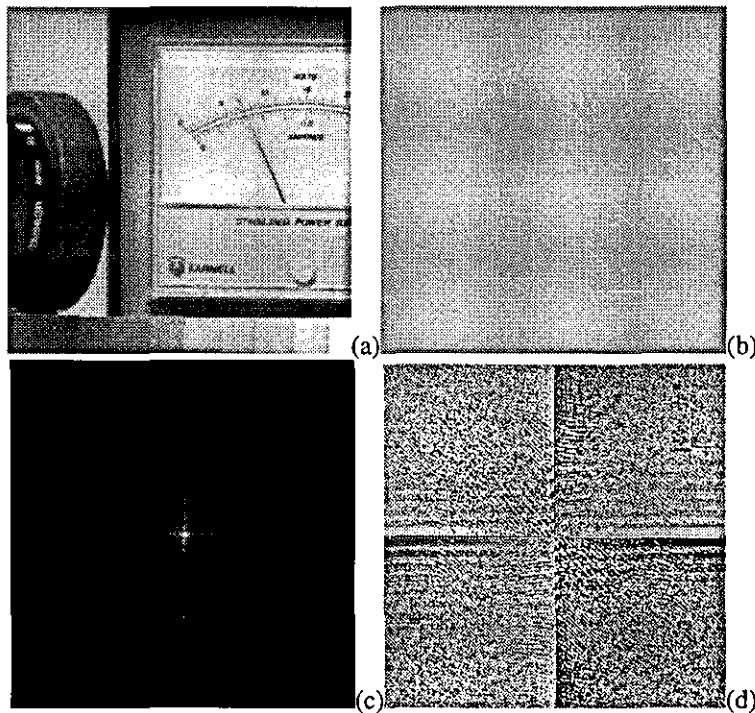


Fig. 4.16 - Original image and its Fourier transform. (a)- image magnitude; (b)- image phase; (c)- Fourier magnitude; (d)- Fourier phase.

Figure 4.17 shows the reconstruction of the original image using the magnitude-only and phase-only synthesis. In Fig. 4.17(a) and (b) is the magnitude-only image (magnitude and phase) which is formed by combining the magnitude of the Fourier transform of the original image in Fig. 4.16(c) with constant phase, in particular with phase zero. While this magnitude-only image reconstruction contains no recognizable features, the phase-only image (magnitude and phase), shown in Fig. 4.17(c) and (d), which is formed by combining the phase of the Fourier transform of the original image in Fig. 4.16(d) with a constant unity magnitude contains many of the features of the original image. An even better phase-only Fourier synthesis can be obtained if the correct Fourier phase is combined with a magnitude which is representative of the class of images of interest. Shown in Fig. 4.18, for example, is the result obtained by combining the phase of the Fourier transform of the original image in Fig. 4.16(d) with

an ensemble average magnitude computed by averaging the Fourier transform magnitudes of several completely different images.

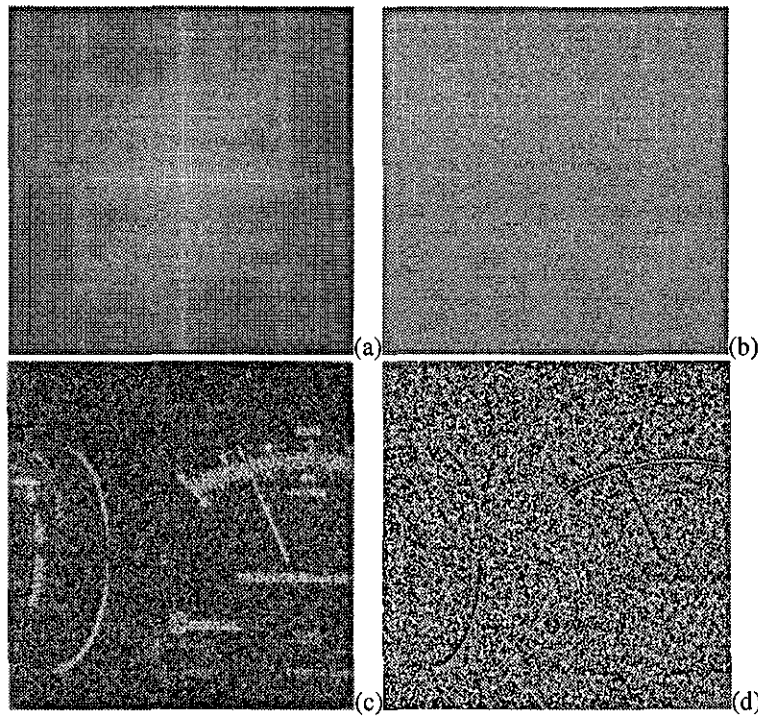


Fig. 4.17 - Fourier synthesis. Magnitude-only: (a)- image magnitude; (b)- image phase;
Phase-only: (c)- image magnitude; (d)- image phase.

From comparison of Fig. 4.17(a), (c) and Fig. 4.18(a) it is clear that the phase-only image, but not the magnitude-only image, retains many of the important features of the original.

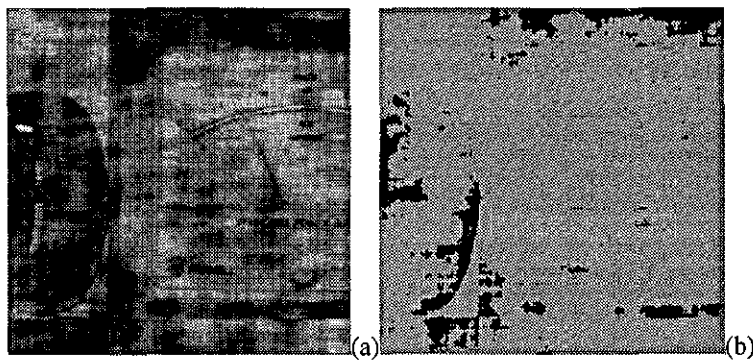


Fig. 4.18 - Fourier synthesis. Phase-only with ensemble magnitude: (a)- image magnitude;
(b)- image phase.

The importance of phase in the representation of signals has also been observed in a number of other applications such as speech processing [28] and acoustical and optical holography [29-35]. In acoustical and optical holography, the hologram corresponds to

the diffraction pattern at a reference plane due to the illumination of an object by a monochromatic source of light. For a two-dimensional object for example, if the reference plane is sufficiently far from the object, the Fraunhofer-type approximation can be made and the spatial diffraction pattern $F(u, v)$ is [36]

$$F(u, v) = \iint_{-\infty}^{+\infty} f(x, y) \exp[-jk(xu + yv)/z] dx dy, \quad (4.4-3)$$

where $f(x, y)$ is the light or sound distribution at the two-dimensional object, x and y are the spatial co-ordinates at the object, u and v are the spatial co-ordinates at the reference plane, $k = 2\pi/\lambda$ is the wave number, λ is the wavelength of the source and z is the distance from the object to the reference plane. From Eq. (4.4-3) one can recognize the diffraction pattern as the two-dimensional Fourier transform of the object with $2\pi u/(\lambda z)$ and $2\pi v/(\lambda z)$ representing the spatial frequencies.

In both the acoustical case and the optical case, the possibility of reconstructing the object from only the phase of the diffraction pattern has been proposed and investigated and in particular it has been demonstrated that reasonable representations of the object can be obtained using this phase synthesis. As mentioned in chapter one, spectral phase is not always available and by contrast no difficulty is encountered in recording the spectral intensity and therefore the spectral magnitude can be calculated. So, the problem now concerns the recovery of the spectral phase information from measurements of the spectral intensity (diffraction pattern) in order to reconstruct the object.

4.4.2 The Gerchberg and Saxton Algorithm

The technique concerning the phase retrieval problem, which is proposed in chapter one, involves Fourier transformations. As mentioned before, for the particular case of coherent illumination the Fourier transform of a coherent wave-front arriving from an object is the Fraunhofer diffraction pattern of the object wave-front. The complete definition of the Fraunhofer diffraction complex function implies the knowledge of the Fourier magnitude and phase of the transformed wave-front. Optically, the Fourier transformation is performed by a lens, and the Fourier magnitude can be obtained from the corresponding spectral intensity, but the phase is either unobservable directly or cannot be determined anywhere as accurately as the intensity. A particularly successful method of solving this problem is the use of the Gerchberg and Saxton [37] and related

algorithms [18-21,38,39]. The recovery of phase information is performed from magnitude information taken from the object and Fourier domains [40,41]. One is interested in reconstructing an optical wave arriving from a coherently illuminated object by taking two intensity measurements, one performed in the spatial object domain or image domain and the other in the spatial frequency domain or Fourier domain, according to the arrangement shown in Fig. 4.19.

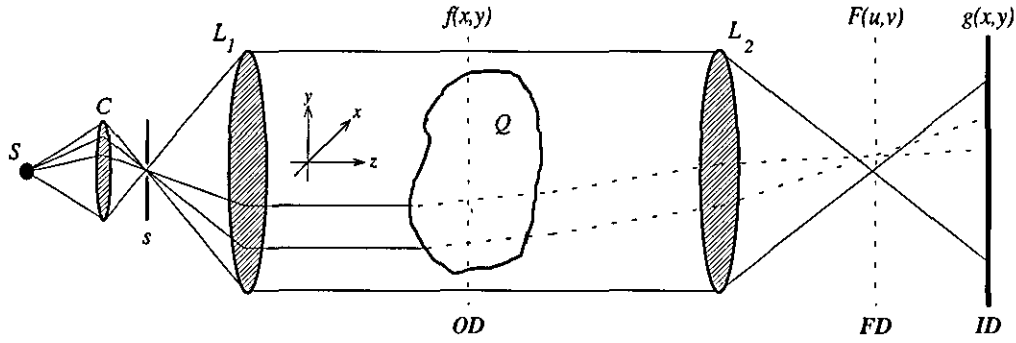


Fig. 4.19 - Optical apparatus for spatial and spectral intensity measurements.

Before describing the algorithm, consider that the optical wave leaving the object disturbance Q is a complex-valued function, $f(x, y)$, denoted by

$$f(x, y) = |f(x, y)| \exp[j\theta(x, y)], \quad (4.4-4)$$

where (x, y) is the two-dimensional pair of spatial co-ordinates, $|f(x, y)|$ is the spatial magnitude and $\theta(x, y)$ is the spatial phase in the object domain OD . Since the object and image domains are conjugate domains, one can define for the image domain ID a wave function $g(x, y)$, denoted by

$$g(x, y) = f(x, y)^* = |f(x, y)| \exp[-j\theta(x, y)], \quad (4.4-5)$$

where $f(x, y)^*$ represents the complex conjugate of Eq. (4.4-4). It should be noted that the measured intensity in the image domain is proportional to the square of the magnitude and will be represented by

$$f_i(x, y) = |f(x, y)|^2, \quad (4.4-6)$$

where $f_i(x, y)$ is the image intensity function.

The object wave front function, $f(x, y)$, is related to its direct Fourier transform (DFT), $F(u, v)$, by

$$F(u, v) = \iint_{-\infty}^{+\infty} f(x, y) \exp[-j2\pi(xu + yv)] dx dy, \quad (4.4-7)$$

where (u, v) is the two-dimensional pair of spatial frequency co-ordinates. $F(u, v)$ is always a complex-valued function, called the Fraunhofer diffraction pattern and denoted by

$$F(u, v) = |F(u, v)| \exp[j\psi(u, v)], \quad (4.4-8)$$

where $|F(u, v)|$ is the spectral magnitude and $\psi(u, v)$ is the spectral phase. The measured intensity in the Fourier domain FD is also proportional to the square of the spectral magnitude and will be represented by

$$F_i(u, v) = |F(u, v)|^2, \quad (4.4-9)$$

where $F_i(u, v)$ is the Fourier intensity function. On the other hand, the diffraction wave-front function, $F(u, v)$, is related to the object wave-front function, $f(x, y)$, by its inverse Fourier transform (IFT) as

$$f(x, y) = \iint_{-\infty}^{+\infty} F(u, v) \exp[j2\pi(xu + yv)] du dv. \quad (4.4-10)$$

In practice; object, image and Fourier intensity functions can be discretized into a sequence of values by taking $N \times N$ equally spaced samples. The importance of this discretization lies in the fact that it allows the use of the computer for information storage and for the development of practical numerical algorithms to process the stored data. The discrete Fourier transform pair that applies to sampled functions, as described in chapter three, is given by

$$F(u, v) = \frac{1}{N^2} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} |f(x, y)| \exp \left[-j \left(\frac{2\pi}{N} (xu + yv) - \theta(x, y) \right) \right], \quad (4.4-11)$$

for $u = 0, 1, 2, \dots, (N-1)$; $v = 0, 1, 2, \dots, (N-1)$, and

$$f(x, y) = \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} |F(u, v)| \exp \left[j \left(\frac{2\pi}{N} (xu + yv) + \psi(u, v) \right) \right], \quad (4.4-12)$$

for $x = 0, 1, 2, \dots, (N-1)$; $y = 0, 1, 2, \dots, (N-1)$, where $|f(x, y)|$ and $|F(u, v)|$ are the magnitude functions calculated by taking the square root of the corresponding sampled

intensity, as indicated in Eqs. (4.4-6) and (4.4-9), and $\theta(x, y)$ and $\psi(u, v)$ are the unknown phase functions in the image and Fourier domains, respectively, that contain most of the information about the object disturbance.

One wishes to recover the unknown phase function, $\theta(x, y)$, or equivalently recover the unknown spectral phase function, $\psi(u, v)$, from measurements of both intensity functions $f_i(x, y)$ and $F_i(u, v)$. An important relation between the measured intensities, which is based on the conservation of energy theorem and referred to as the generalized Parseval-Rayleigh theorem, can be defined for the discrete case, as discussed in chapter three, by

$$\sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f_i(x, y) = N^2 \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} F_i(u, v), \quad (4.4-13)$$

where $f_i(x, y)$ and $F_i(u, v)$ are the measured intensity functions. This relation will be useful to eventually modify the measured data in order to satisfy the above theorem and therefore to improve the results in the reconstruction process.

Since intensity recordings of sampled wave-fronts can be made conveniently in both image and Fourier domains, one can describe a method for determining the complete wave-front function (magnitude and phase) in one domain from such intensity records. As mentioned before, the method is based on discrete Fourier transforms and depends on there being a Fourier transform relation between the waves in the two domains. The basis is the Gerchberg and Saxton algorithm that concerns an iterative procedure involving discrete Fourier transformation back and forth between the image and Fourier domains and application of the recorded intensities in each domain. The direct and inverse Fourier transforms are done by means of the fast Fourier transform algorithm, FFT and IFFT, on the computer. The algorithm consists of the following simple steps:

- Generate an estimate of the complex object wave-front $h(x, y)$, and compute the FFT to obtain the corresponding complex diffraction pattern $H(u, v)$;
- Replace the magnitude of the computed Fourier transform with the calculated magnitude, $|F(u, v)| = F_i(u, v)^{1/2}$, to form an estimate of the complex diffraction pattern $H'(u, v)$;
- Compute the IFFT of the estimate, $H'(u, v)$, to obtain the corresponding complex object wave-front $h'(x, y)$;
- Replace the magnitude of the computed Fourier transform with the calculated magnitude, $|f(x, y)| = f_i(x, y)^{1/2}$, to form a new estimate of the complex object wave-front $h(x, y)$.

In terms of equations this is, for the k th iteration:

$$H_k(u, v) = |H_k(u, v)| \exp[j\phi_k(u, v)], \quad (4.4-14)$$

$$H'_k(u, v) = |F(u, v)| \exp[j\phi_k(u, v)], \quad (4.4-15)$$

$$h'_k(x, y) = |h'_k(x, y)| \exp[j\phi'_k(x, y)], \quad (4.4-16)$$

$$\begin{aligned} h_{k+1}(x, y) &= |f(x, y)| \exp[j\phi'_k(x, y)] \\ &= |f(x, y)| \exp[j\phi_{k+1}(x, y)] \end{aligned} \quad (4.4-17)$$

where $h_k(x, y)$, $\phi_k(x, y)$, $H'_k(u, v)$, and $\phi_k(u, v)$ are estimates of $f(x, y)$, $\theta(x, y)$, $F(u, v)$, and $\psi(u, v)$, respectively.

Notice that the initial estimate, $h_0(x, y)$, is obtained by combining the magnitude, calculated from the recorded intensity, with a random phase function, $\phi_0(x, y)$, with values uniformly distributed between π and $-\pi$. It is also possible to start a similar iterative process from an initial estimate, $H'_0(u, v)$, of the complex diffraction pattern instead of an estimate of the complex object wave-front. Concerning the initial phase function, it can be convenient in some situations to use instead of a random phase distribution another phase distribution which represents a phase guess for the problem, that has previously been obtained in some way from the information available.

This algorithm, as depicted in Eqs. (4.4-14) through (4.4-17), is easily generalized to a very large class of problems and it can be used for any problem in which partial constraints (in the form of measured data or information known a priori) are known in each of the two domains. The operation of the algorithm is schematically described on the flow chart shown in Fig. 4.20. The algorithm simply transforms back and forth between the two domains, satisfying the constraints in each domain before returning to the other domain.

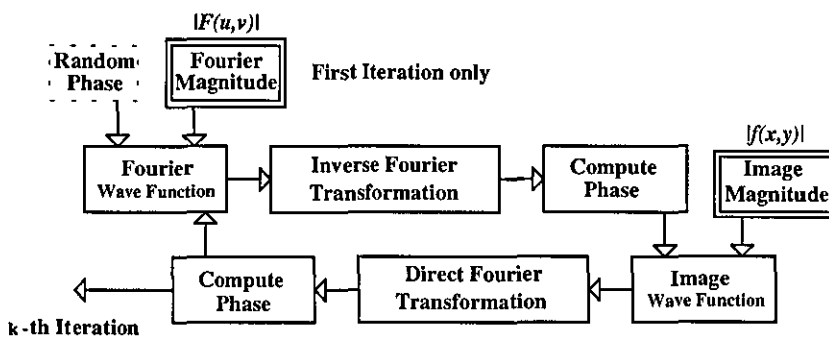


Fig. 4.20 - The Gerchberg and Saxton iterative algorithm.

The input data to the algorithm are the magnitudes calculated from the sampled intensities in both the image and Fourier domains. These two sets of data are accessed

once per complete iteration. To begin, as mentioned above, a uniform random number generator or a phase guess is used to generate an array of numbers which serves as the initial estimate of the phase in the Fourier domain. This array is then combined with the corresponding Fourier magnitude data and the inverse Fourier transform IFT of this synthesized complex function is evaluated. The phase of the resulting complex image function is calculated and combined with the corresponding image magnitude data. This new function is then transformed by direct Fourier transform DFT. The phase of the resulting complex Fourier function is computed and combined with the Fourier magnitude data to form a new estimate. This process should be repeated in order to obtain a better estimate for the complex diffraction pattern function.

The convergence of the algorithm can be monitored by computing the squared error in each iteration. In the Fourier domain the square error is the sum of the squares of the amounts by which $H_k(u, v)$, the computed Fourier transform, violates the Fourier domain constraints. The squared error for the k th iteration, E_{Fk}^2 , can be expressed by

$$E_{Fk}^2 = \frac{1}{N^2} \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} [|H_k(u, v)| - |F(u, v)|]^2. \quad (4.4-18)$$

Similarly, in the image domain, the squared error is the sum of the squares of the amounts by which $h_k'(x, y)$, the computed inverse Fourier transform, violates the image domain constraints. The square error for the k th iteration, E_{Ik}^2 , can be expressed by

$$E_{Ik}^2 = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} [|f(x, y)| - |h_k'(x, y)|]^2. \quad (4.4-19)$$

This generalization of the Gerchberg and Saxton algorithm is usually referred to as the error reduction algorithm since the error decreases (or at least stays the same) at each iteration [42]. Depending on the starting point, during the iterative process to retrieve the phase functions, one or both the Eqs. (4.4-18) and (4.4-19) should be used to calculate the square error at each iteration. When the squared error achieves zero or a pre-defined value then the correct solution or an approximation to the correct solution has been found. Instead of the squared error the normalized root mean square (rms) error is often used. In the Fourier domain, the normalized rms error is the square root of E_{Fk}^2 divided by the constant B_F^2 , where

$$B_F^2 = \frac{1}{N^2} \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} |F(u, v)|^2 = \frac{1}{N^2} \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} F_i(u, v). \quad (4.4-20)$$

Similarly, the normalized rms error in the image domain is the square root of E_{rk}^2 divided by the constant B_I^2 , where

$$B_I^2 = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} |f(x, y)|^2 = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f_i(x, y). \quad (4.4-21)$$

The described algorithm has been implemented and integrated on the computer application program discussed in chapter two. Due to computational restrictions, the phase recovery process only recovers relative phase distributions and not absolute phase distributions. This seems to be a limitation but in reality it is not. The implemented algorithm uses a computer mathematical function called *arctan* to compute the argument of a complex number between $-\pi$ and $+\pi$. Because of this inverse tangent operation, the phase distribution is wrapped into the range $[-\pi, +\pi]$ and 2π jumps occur for phase variations of more than 2π . The absolute phase distribution can be obtained by adequate processing of the relative phase distribution with specific phase unwrapping algorithms [43-47]. In particular, the algorithm developed by Ghiglia [43], which is based in fast cosine transforms, will be used to obtain when necessary all the unwrapped phase distributions.

In the next chapters, computer simulated object wave-fronts and real experimental object wave-fronts will be used as source to produce information (intensity and magnitude) to iteratively retrieve phase distributions as it has been described.

4.5 Summary

In this chapter the concepts of image formation and image reconstruction have been presented and described. The emphasis placed on the interference and diffraction effects reflect their importance as an intermediate step in the image formation process. Due to the particular interest in Fraunhofer diffraction and coherent illumination throughout this thesis a detailed discussion focusing on these concepts has been made and several illustrating examples have been included with the corresponding theoretical basis.

Concerning the image formation process two models of image formation have been described and special attention has been devoted to coherent imaging and its practical implementation.

Supported by the Fourier optics theory, the importance of phase in the image reconstruction process has been demonstrated and an example has been presented for

illustration purposes. Since phase information plays an important role in the image formation and reconstruction processes, and phase information is usually not available in an optical system, an algorithm for phase retrieval has been discussed. The algorithm is basically the Gerchberg and Saxton iterative algorithm. It recovers phase information from intensity measurements made in both the image and diffraction planes in an optical system. The discrete operation of the algorithm has been fully described and equations to monitor the convergence of the iterative procedure have been deduced.

References

1. F. A. Jenkins and H. E. White, *Fundamentals of Optics*, 4th Edition, McGraw-Hill Ed., Singapore (1981).
2. E. Hecht, *Optics*, 2nd Edition, Addison-Wesley Publishing Company, Inc., (1987).
3. R. D. Guenther, *Modern Optics*, John Wiley & Sons, Inc., New York (1990).
4. S. G. Lipson and H. Lipson, *Optical Physics*, Cambridge University Press, London (1969).
5. J. W. Goodman, *Introduction to Fourier Optics*, Reissued Edition, McGraw-Hill, Inc., New York (1988).
6. R. W. Wood, *Physical Optics*, 2nd Edition, Dover Publications, Inc., New York (1968).
7. M. Born and E. Wolf, *Principles of Optics - Electromagnetic Theory of Propagation, Interference and Diffraction of Light*, 2nd Edition, Pergamon Press, Oxford (1964).
8. G. N. Watson, *A Treatise on the Theory of Bessel Functions*, Cambridge University Press, London (1922).
9. E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis*, 4th Edition, Cambridge University Press, London (1952).
10. G. B. Airy, "On the Diffraction of an Object-Glass with Circular Aperture", *Trans. Cambridge Phil. Soc.* **5**, 283-291 (1835).
11. H. Lipson, *Optical Transforms*, Academic Press, Inc., London (1972).
12. E. G. Steward, *Fourier Optics: an Introduction*, 2nd Edition, Ellis Horwood Limited, Chichester (1989).

13. E. Abbe, "Beiträge zur Theorie des Mikroskops und der Mikroskopischen Wahrnehmung", *Archiv. f. Microskop. Anat.* **9**, 413-468 (1873). (*Collected Works* **1** (1904))
14. A. B. Porter, "On the Diffraction Theory of Microscopic Vision", *Phil. Mag.* **11**, 154-166 (1906).
15. L. Rayleigh, "On the Theory of Optical Images, with Special Reference to the Microscope", *Phil. Mag.* **42**, 167-195 (1896).
16. P. M. Duffieux, *The Fourier Transform and its Applications to Optics*, 2nd Edition, John Wiley & Sons, Inc., New York (1983).
17. W. O. Saxton, *Computer Techniques for Image Processing in Electron Microscopy*, Academic Press, Inc., New York (1978).
18. R. A. Gonsalves, "Phase Retrieval from Modulus Data", *J. Opt. Soc. Am.* **66**, 961-964 (1976).
19. J. R. Fienup, "Reconstruction of an Object from the Modulus of Its Fourier Transform", *Optics Letters* **3**, 27-29 (1978).
20. W. Kim and M. H. Hayes, "Phase Retrieval using Two Fourier-Transform Intensities", *J. Opt. Soc. Am. A* **7**, 441-449 (1990).
21. J. C. Dainty and J. R. Fienup, "Phase Retrieval and Image Reconstruction for Astronomy", in *Image Recovery: Theory and Application* (H. Stark Ed., Academic Press, Inc., San Diego) **7**, 231-275 (1987).
22. G. Ramachandran and R. Srinivasan, *Fourier Methods in Crystallography*, John Wiley & Sons, Inc., New York (1970).
23. L. S. Taylor, "The Phase Retrieval Problem", *IEEE Trans. on Antennas & Propagation* **AP-29**, 386-391 (1981).
24. H. A. Ferwerda, "The Phase Reconstruction Problem for Wave Amplitudes and Coherence Functions", in *Inverse Source Problem in Optics* (H. P. Baltes Ed., Springer-Verlag, Berlin) **2**, (1978).

25. A. V. Oppenheim and J. S. Lim, "The Importance of Phase in Signals", *Proc. of the IEEE* **69**, 529-541 (1981).
26. T. G. Stockham, T. M. Cannon, and R. B. Ingebretson, "Blind Deconvolution through Digital Signal Processing", *Proc. of the IEEE* **63**, 678-692 (1975).
27. R. Srinivasan, "The Significance of the Phase Synthesis", *Proc. of the Indian Acad. Sci.* **53-A**, 252-261 (1961).
28. A. V. Oppenheim, J. S. Lim, G. Kopek, and S. C. Pohlig, "Phase in Speech and Pictures", *Proc. of the IEEE International Conference on Acoustic, Speech, and Signal Processing*, 632-637 (1979).
29. L. Larimore, H. M. A. El-Sum, and A. F. Metherell, "Acoustical Holograms using Phase Information only", *Applied Optics* **8**, 1533-1536 (1969).
30. A. F. Metherell, "The Relative Importance of Phase and Amplitude in Acoustical Holography", in *Acoustical Holography* (A. F. Metherell and L. Larimore Ed., Plenum Press, Inc., New York) **2**, (1970).
31. J. Power, J. Landry, and G. Wade, "Computed Reconstruction from Phase-Only and Amplitude-Only Holograms", in *Acoustical Holography* (A. F. Metherell and L. Larimore Ed., Plenum Press, Inc., New York) **2**, (1970).
32. D. Kermisch, "Image Reconstruction from Phase Information only", *J. Opt. Soc. Am.* **60**, 15-17 (1970).
33. T. Yatagai and M. Takeda, "Effect of Phase Nonlinearity in Kinoform", *Optik* **43**, 337-352 (1975).
34. M. Ahmed, K. Y. Wang, and A. F. Metherell, "Holography and its Applications to Acoustic Imaging", *Proc. of the IEEE* **67**, 466-483 (1979).
35. R. D. Baglai, "Numerical Image Reconstruction from Kinoforms and Distorted Holograms", *Trans. Automatic Monitoring Measuring* **2**, 15-24 (1977).
36. T. S. Huang, "Digital Holography", *Proc. of the IEEE* **59**, 1335-1346 (1971).

37. R. W. Gerchberg and W. O. Saxton, "A Practical Algorithm for the Determination of Phase from Image and Diffraction Plane Pictures", *Optik* **35**, 237-246 (1972).
38. G. Lu, Z. Zhang, F. T. S. Yu, and A. Tanone, "Pendulum Iterative Algorithm for Phase Retrieval from Modulus Data", *Optical Engineering* **33**, 548-555 (1994).
39. G. Yang, B. Dong, B. Gu, *et. al.*, "Gerchberg-Saxton and Yang-Gu algorithms for Phase Retrieval in a Nonunitary Transform System: a Comparison", *Applied Optics* **33**, 209-218 (1994).
40. A. M. J. Huizer and H. A. Ferwerda, "On the Problem of Phase Retrieval in Electron Microscopy from Image and Diffraction Pattern", *Optik* **46**, 407-420 (1976).
41. R. W. Gerchberg and W. O. Saxton, "Phase Determination from Image and Diffraction Plane Pictures in the Electron Microscope", *Optik* **34**, 275-284 (1971).
42. J. R. Fienup, "Comparison of Phase Retrieval Algorithms", in *Advances in Computer Vision and Image Processing* (T. S. Huang Ed., Jai Press, Inc., London) **4**, 191-225 (1984).
43. D. C. Ghiglia and L. A. Romero, "Robust Two-dimensional Weighted and Unweighted Phase Unwrapping that uses Fast Transforms and Iterative Methods", *J. Opt. Soc. Am. A* **11**, 107-117 (1994).
44. D. C. Ghiglia, G. A. Mastin, and L. A. Romero, "Cellular Automata Method for Phase Unwrapping", *J. Opt. Soc. Am. A* **35**, 267-280 (1987).
45. J. M. Huntley and H. Saldner, "Temporal Phase-Unwrapping Algorithm for Automated Interferogram Analysis", *Applied Optics* **32**, 3047-3052 (1993).
46. J. M. Huntley, "Noise-Immune Phase unwrapping Algorithm", *Applied Optics* **28**, 3268-3270 (1989).
47. T. R. Judge, C. Quan, and P. J. Bryanston-Cross, "Holographic Deformation Measurements by Fourier Transform Technique with Automatic Phase Unwrapping", *Optical Engineering* **31**, 533-543 (1992).

CHAPTER 5

Phase Retrieval for Wave-Front Reconstruction

5.1 Introduction

In the previous chapter a numerical algorithm for the phase retrieval problem has been presented and described. This chapter is now devoted to the application of such an algorithm to recover the phase distributions associated to specific real optical object disturbances. With the recovered phase information it will be shown that it is possible to reconstruct the corresponding complex wave-front function in terms of a discrete complex function recreated on the computer. Consequently, the optical object disturbance characterization can be performed on the computer using the recreated discrete complex function.

First, the algorithm will be used with a set of computer simulated optical object disturbances, and finally it will be used with a set of real optical object disturbances. The complex wave-front reconstruction of an optical object disturbance implies the knowledge of both the magnitude and phase distributions of the object. Magnitude information can easily be obtained from measurements of intensity. The phase, which is either unobservable directly or can not be determined anywhere as the intensity, will be recovered from single measurements of intensity using the above mentioned algorithm.

In the next sections the computer simulated results and the real experimental results will be presented and discussed in order to establish the capabilities of this phase retrieval procedure as well as the practical limitations of its implementation.

5.2 Methodology

The algorithm is an iterative procedure based on the Gerchberg and Saxton algorithm [1], which has been implemented and included in the computer application program discussed in chapter two. The algorithm recovers the phase distribution using the magnitude information in the image and Fourier domains associated with an optical object disturbance. For each iteration the computer executes a direct and inverse Fourier transform pair using the discrete complex functions obtained from the magnitude information for both the image and Fourier domains. These two magnitude distributions

that have been obtained from the intensity measurements are the constraints that will be imposed in each domain for each iteration. The algorithm iteratively recovers the phase distributions for both the image and Fourier domains which agree through the Fourier transformation with the imposed constraints.

The iterative procedure considered here always starts by generating a random phase distribution for the Fourier domain which is uniformly distributed in the interval $[-\pi, +\pi]$, as described in chapter four Section 4.4.2. It is possible to start the iteration process by using a guess for the phase distribution instead of the random phase if this phase guess can somehow be made available. This phase information is then combined with the corresponding magnitude information available for the Fourier domain to recreate an estimate of the discrete complex function associated with a particular object disturbance which is called the complex diffraction pattern function of the object.

The inverse Fourier transform of this complex function is another discrete complex function which represents the complex image function of the object. The phase distribution is calculated and combined with the corresponding magnitude information available for the image domain to recreate also an estimate of the discrete complex image function. Now, the direct Fourier transform is performed to obtain a new estimate of the complex diffraction pattern function. The iteration cycle ends when the new phase distribution is calculated and the normalized squared error in the current iteration has also been calculated. This cycle of iterations should be executed several times in order to obtain a better and close estimate of the phase distribution for the Fourier domain. The convergence of the algorithm can be monitored through the normalized squared error.

As mentioned in chapter four, the phase recovery process retrieves the relative phase distribution and not the absolute phase distribution. The relative phase is the absolute phase wrapped into the range $[-\pi, +\pi]$ and 2π jumps occur for phase variations of more than 2π . For this reason the relative or wrapped phase distribution retrieved by the algorithm should be unwrapped to obtain the absolute phase distribution using phase unwrapping algorithms such as the one mentioned in reference [2]. It is possible that different relative phase distributions can be recovered by the algorithm for the same object disturbance, but they will represent the same absolute phase distribution. The unwrapping process should always be performed to evaluate the recovered phase distributions. The relative phase distribution coincides with the absolute phase distribution when the phase changes are confined in the range $[-\pi, +\pi]$. The algorithm can fail in the recovery process. A failure in the retrieval process occurs when the

recovered phase distributions are not consistent for different iteration processes executed in the same conditions. It is convenient to execute several iteration processes before inferring about its convergence to the correct solution. The conditions for the success of the phase retrieval will be discussed and illustrated in the next section.

In the following sections, unless otherwise stated, intensity and magnitude distributions will be displayed as pictures where the displayed information corresponds to the equivalent distribution normalized in the interval $[0,1]$. The value 0 appears displayed as a black level and the value 1 as a white level. The intermediate values appear as different grey levels. The relative phase distributions will also be displayed as normalized distributions in the interval $[-\pi, +\pi]$. For this case, the value $-\pi$ appears displayed as a black level and the value $+\pi$ as a white level. Different grey levels are associated with intermediate values in the considered interval. The absolute phase distributions will be displayed in the same way as the intensity and magnitude distributions, but in this case additional information concerning the scale will be given in the supporting text.

5.3 Computer Simulations

The research that has been carried out concerns the phase retrieval and therefore the reconstruction of complex wave-fronts for simple simulated object disturbances. The simulated objects that will be used throughout this section are close representations of the equivalent real object disturbances in terms of optical properties. The application of the iterative algorithm to recover the phase distributions will use the intensity information simulated on the computer for both the image and Fourier domains. Later, the obtained results will be extended to real experimental situations. Emphasis is placed on phase-object disturbances, that is, object disturbances that do not modify the magnitude distribution of an optical wave-front that travels through it, but do modify its phase distribution. By contrast, magnitude-object disturbances are disturbances that do modify the optical wave-front magnitude distribution, but do not modify its phase distribution. In general, a reference to an object disturbance means that both the magnitude and phase distributions of an optical wave-front could be modified by the object.

It should be mentioned that before starting the computer simulations a compromise has to be made between the scale and size of both the image and Fourier intensity information. As well-known from the Fourier theory [3], the scale in the Fourier domain

is inversely proportional to the scale in the object or image domain (see also Sections 3.3 and 3.4 in chapter three). This suggests that the Fourier domain information must be simulated with a magnification inversely proportional to the magnification used in the image domain. For convenience the array size in all the simulations will be 512×512 pixel array for both the image and Fourier domains. The image domain information will be de-magnified 10x in order to obtain, in a 512×512 array, reasonably detailed Fourier diffraction pattern information. The value 10x was chosen for some other reasons, which will be explained and discussed later in this chapter, in order to obtain a correct agreement between the simulations and the experiment described in Section 5.4.

5.3.1 Intensity and Magnitude Measurements

As mentioned in the previous section the simulated information for the image and Fourier domains will be stored in 512×512 pixel arrays. In particular, for the image domain the simulated intensity distribution, that will be used in the following simulations, is a centred circular aperture 480 pixels in diameter filled with a truncated Gaussian distribution of amplitudes, as shown in Fig. 5.1(a) and the horizontal profile along the line in (b). The equivalent simulated magnitude distribution, which has been calculated by taking the square root of the intensity distribution, is also shown in Fig. 5.1(c) and the corresponding profile in (d).

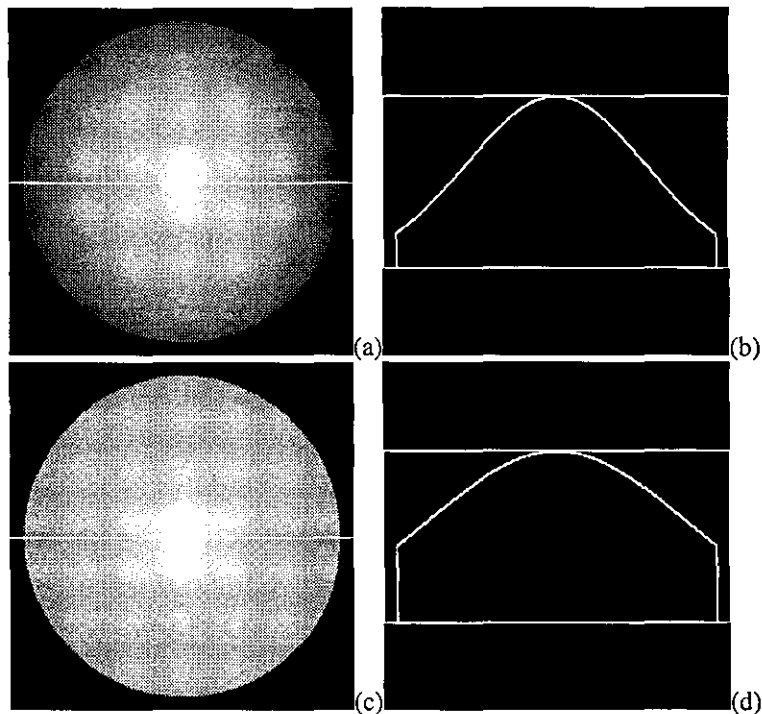


Fig. 5.1 - The image domain information. (a)- intensity distribution; (b)- profile marked in (a); (c)- magnitude distribution; (d)- profile marked in (c).

Assuming that the phase distribution associated with the above magnitude distribution is constant and equal to zero over the considered aperture, one can use the direct fast Fourier transform algorithm to calculate the corresponding diffraction pattern distribution in the Fourier domain. Figure 5.2 shows respectively in (a) the intensity distribution of the calculated diffraction pattern and in (b) the indicated horizontal profile.

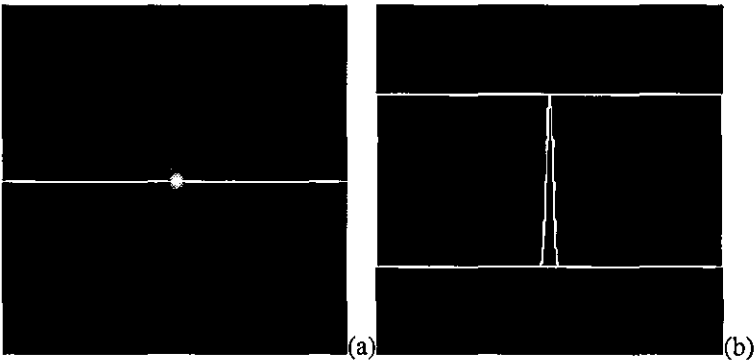


Fig. 5.2 - The Fourier domain information. (a)- intensity distribution; (b)- profile marked in (a).

The diffraction pattern magnitude distribution can be obtained directly from the result of the Fourier transformation which is shown in Fig. 5.3(a) with the corresponding profile in (b). It is now important to observe the resulting magnitude distribution obtained when the square root of the intensity distribution is performed, as shown in Fig. 5.3(c) and (d).

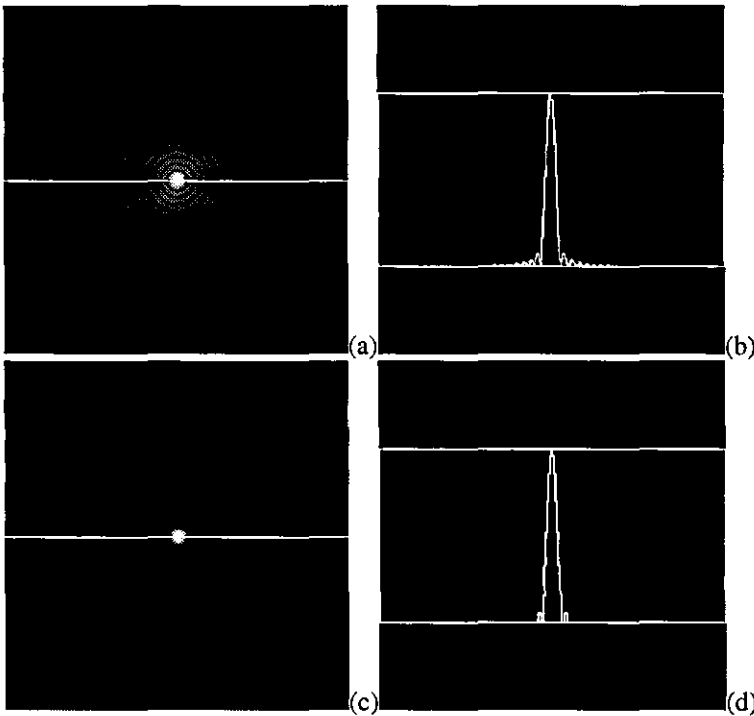


Fig. 5.3 - The Fourier domain information. (a)- magnitude distribution; (b)- profile marked in (a); (c)- calculated magnitude distribution from intensity; (d)- profile marked in (c).

One can see that the square root operation has distorted considerably the contents of the diffraction pattern magnitude if compared with the correct magnitude information. The reason is related to the lack of information in the intensity distribution and consequently all the detail and information in the high frequency components is completely lost. This is a real problem when only intensity measurements are available in the majority of practical applications it being impossible to obtain from them the equivalent magnitude measurements.

The problem is only noticeable with the information in the Fourier domain due to the low level information contents in the high frequency components. The calculation of the magnitude distribution from the intensity distribution in the image plane does not reveal this problem. Consequently, this loss of information in the Fourier diffraction pattern can affect the phase retrieval process, as will be shown in the next section. If the phase retrieval process fails due to losses of information in the calculated magnitude from its computer simulated intensity distribution, it probably fails if the information is taken from a real practical application where only intensity measurements can normally be made. This means that accurate measurements of the intensity in the Fourier domain must be achieved in order to overcome or at least reduce the loss of information, as will be discussed in the next sections.

5.3.2 Simulation Results

The described iterative algorithm will be used to retrieve the phase distribution of a simulated object wave-front. The input data to the iterative algorithm is obtained from computer simulated functions for the image and Fourier domains.

The simulation results are divided into two different sets of the phase retrieval process. For the first set of simulations, one function concerns the distribution of the simulated intensity information in the image domain, the image of the object disturbance shown in Fig. 5.1(a), and the other function the distribution of the simulated intensity information in the Fourier domain, the diffraction pattern of the corresponding object, which are then used to calculate the corresponding magnitude distribution functions. For the second simulation set, one function concerns the distribution of the simulated magnitude information in the image domain, the image of the object disturbance shown in Fig. 5.1(b), and the other function the distribution of the simulated magnitude information in the Fourier domain, the diffraction pattern of the corresponding object. The simulated data is stored on the computer after the necessary 10x de-magnification of the

magnitude distribution in the image plane for scale agreement purposes. After these operations, these two sets of magnitude functions are the effective input functions to the algorithm.

The algorithm iteratively estimates the phase distribution function in one of the selected domains. For this case in particular, it will be always select the image domain. The computer for each iteration executes one direct Fourier transformation and inverse Fourier transformation pair over the input data. The iterative procedure considered here always starts by generating a random phase function for the Fourier domain which is uniformly distributed in the interval $[-\pi, +\pi]$. Since the algorithm is an iterative procedure the convergence of the algorithm will be monitored through the calculated normalized root mean squared error. The number of iterations executed will vary for different retrieval processes.

The following text concerns the presentation and discussion of the simulation results obtained with the stored data in the two different sets of phase retrieval for three phase-object disturbances. The simulated phase-objects are respectively a Fresnel bi-prism phase disturbance, an exponential phase disturbance, and a ramp discontinuity phase disturbance, to cover both linear and non-linear phase change distributions.

Fresnel Bi-Prism Phase Disturbance

Figure 5.4(a) depicts the simulated phase distribution of a Fresnel bi-prism viewed through a circular aperture 480 pixels in diameter. The phase change is a linear phase change that increases from the left and right-hand sides to the middle of the aperture. The phase change from one side to the middle corresponds to a change from -3π to $+3\pi$, 6π rad being the total simulated phase change over the whole aperture, as shown by the profile in Fig. 5.4(b).

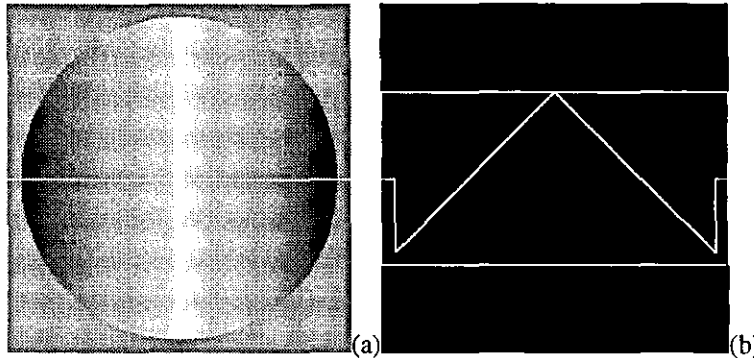


Fig. 5.4 - The simulated phase change for the Fresnel bi-prism. (a)- phase distribution; (b)- profile marked in (a).

This phase distribution was combined with the simulated magnitude distribution (see Fig. 5.1(c)) to create a complex image function of the considered object. The fast Fourier transform of this function was then calculated in order to obtain the corresponding simulated complex diffraction pattern function. Figure 5.5 shows in (a) the diffraction pattern intensity distribution and in (b) the indicated profile. Also, in Fig. 5.5(c) and (d) is shown the corresponding simulated magnitude distribution.

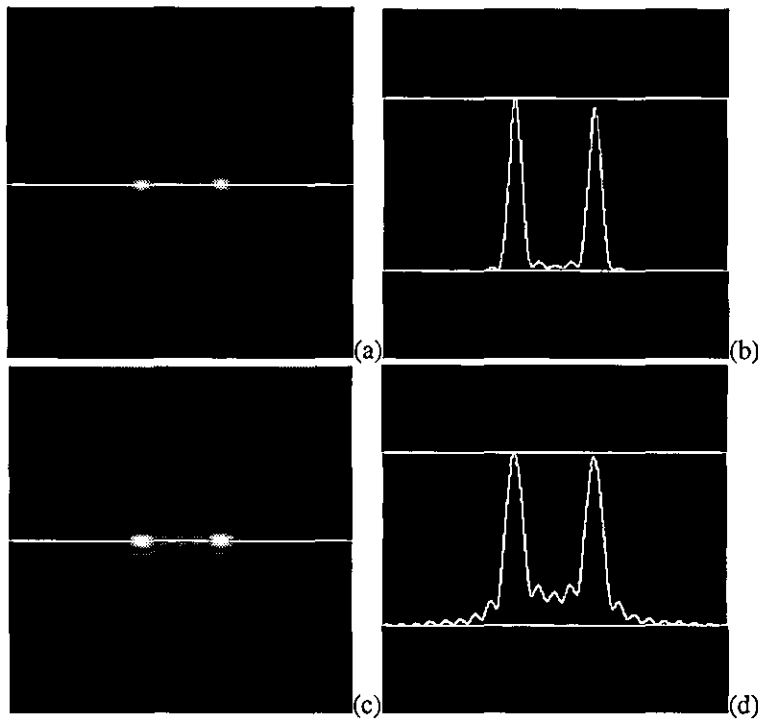


Fig. 5.5 - The diffraction pattern for the Fresnel bi-prism. (a)- intensity distribution; (b)- profile marked in (a); (c)- magnitude distribution; (d)- profile marked in (c).

It should be mentioned that a linear phase change in the object or image domain corresponds to a linear shift of the diffraction pattern in the Fourier domain. As can be seen in Fig. 5.5, due to the phase distribution the diffraction pattern of the single circular aperture was split into two similar diffraction patterns which were linearly shifted from the central region along the horizontal direction. The distance between the two central maxima is 120 pixels and the individual shift due to the symmetry of the phase change is 60 pixels. Since the phase change is linear and the shift in the Fourier domain is also linear, from this result one can obtain the relation between the phase change and the shift of the diffraction pattern which gives the sensitivity and resolution of the measurements that can be made with the Fourier diffraction information. The relation obtained is that for a phase change of 2π the diffraction pattern is shifted 20 pixels and therefore the minimum phase change that can be detected is $2\pi/20$ rad. This explains in fact that a limited phase change can be measured from the diffraction pattern, when assuming a linear phase distribution in the object or image domain.

Since the arrays to store the information have a limited size, 512×512 pixels, it seems possible to measure approximately 25 phase changes with an amount of 2π rad each. But for this case, one can expect that part of the diffraction pattern goes out of the array limits and consequently it will be cut somewhere. This problem can in principle be avoided if larger arrays, say 1024×1024 pixels or more, are used instead. Increasing the size of the arrays has the disadvantage of the increasing processing time and a special computer configuration to manipulate large arrays will be necessary. Keeping the array sizes and the scale relations for the image and Fourier domains as fixed before, one can safely measure in these conditions approximately 12 linear phase changes of 2π rad.

The objective is now to use these simulated magnitude functions to recover the original phase distribution. The first set of procedures to recover the phase is performed using the magnitude distributions for the image and Fourier domains that have been calculated from the corresponding intensity distributions, respectively shown in Fig. 5.1(a) and Fig. 5.5(a). This set was composed by repeating a 200 iteration process several times feeding the algorithm with the magnitude functions mentioned previously.

All the results obtained in this set were not consistent with each other. Figure 5.6(a) shows one of these recovered wrapped phase distributions when the normalized root mean squared error was 2.38×10^{-1} after 200 iterations.

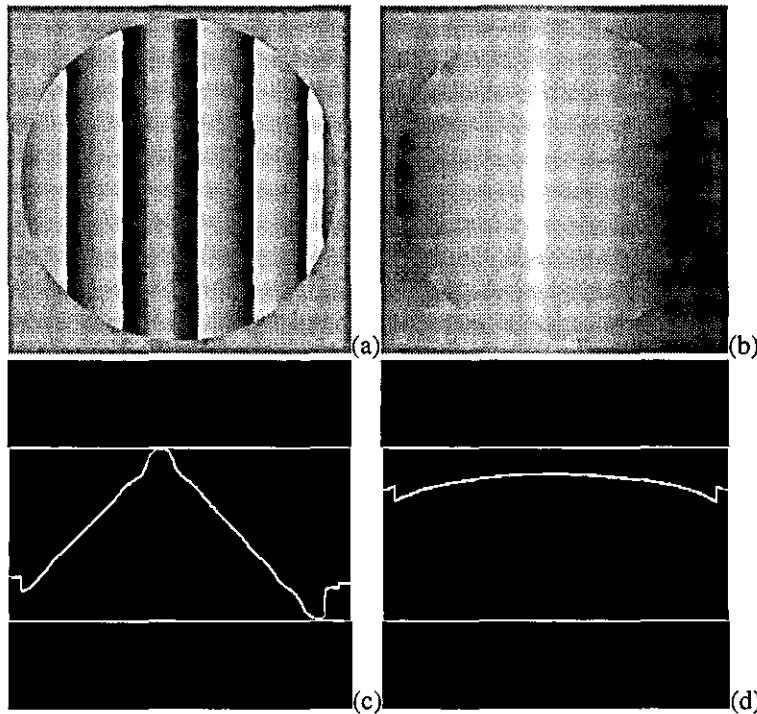


Fig. 5.6 - Recovered phase function for the Fresnel bi-prism. (a)- relative distribution; (b)- absolute distribution; (c)- middle horizontal profile of (b); (d)- middle vertical profile of (b).

After 200 more iterations the error did not change from the previous value which was 2.38×10^{-1} . The corresponding unwrapped phase distribution is shown in Fig. 5.6(b). Also, in Fig. 5.6(c) and (d) are shown respectively the middle horizontal and vertical profiles of the unwrapped phase.

As one can observe from the previous pictures and profiles, the recovered phase distribution does not agree with the original phase distribution and one must consider this set of procedures to retrieve the phase unsuccessful. The reason for the failure has already been mentioned before and is totally related with the loss of vital information in the high frequency components of the diffraction pattern function. Also, the normalized root mean squared error appears to be stagnated which suggests that the convergence of the algorithm has stopped and it is not worth executing more iterations.

The second set of procedures to recover the phase is now performed using the magnitude distributions for the image and Fourier domains that have been simulated, shown in Fig. 5.1(c) and Fig. 5.5(c) respectively. This set was also composed by repeating a 200 iteration process several times, as it has been done before.

All the results obtained in this set were quite consistent with each other. Figure 5.7(a) shows one of these recovered wrapped phase distributions when the normalized root mean squared error was 4.51×10^{-4} after 200 iterations.

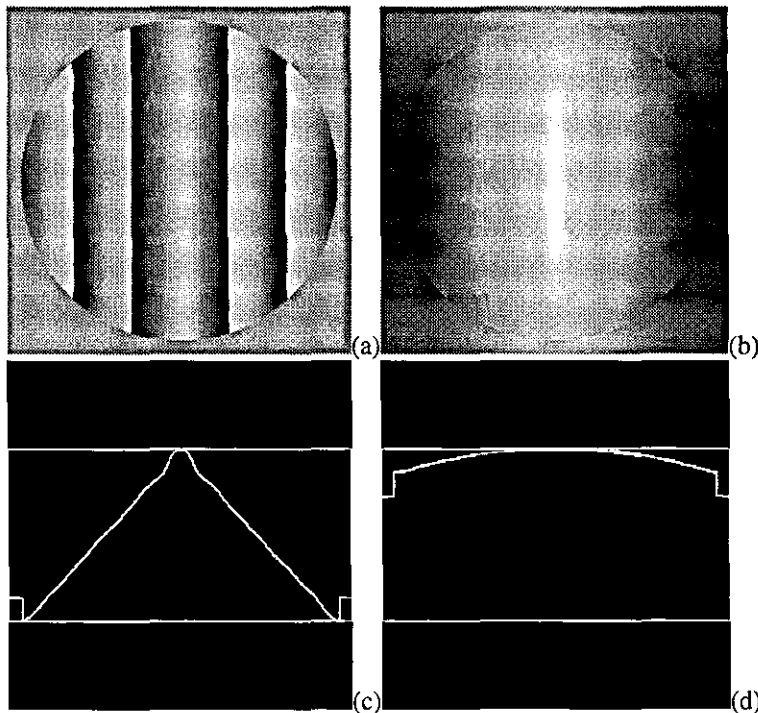


Fig. 5.7 - Recovered phase function for the Fresnel bi-prism. (a)- relative distribution; (b)- absolute distribution; (c)- middle horizontal profile of (b); (d)- middle vertical profile of (b).

After 200 more iterations the error has been reduced to 8.82×10^{-6} . The corresponding unwrapped phase distribution is shown in Fig. 5.7(b). Also, in Fig. 5.7(c) and (d) are shown respectively the middle horizontal and vertical profiles of the unwrapped phase.

The results obtained in this set serve to prove that the algorithm converges to the correct solution when correct and accurate information is used in the iterative process. The normalized root mean squared error shows clearly the convergence of the algorithm. The reproducibility and consistency of the results in all the different processes executed suggests the success of the phase retrieval algorithm to recover the original phase distribution.

Exponential Phase Disturbance

Figure 5.8(a) depicts the simulated phase distribution of an exponential disturbance viewed through a circular aperture 480 pixels in diameter. In this case, the phase change is a non-linear phase change that increases from the top to the bottom of the aperture. The phase change along the vertical direction corresponds to a change from -3π to $+3\pi$, 6π rad being the total simulated phase change over the whole aperture, as shown by the profile in Fig. 5.4(b).

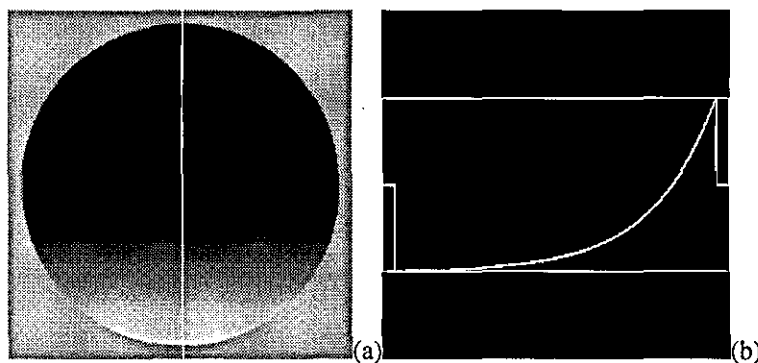


Fig. 5.8 - The simulated phase change for the exponential disturbance. (a)- phase distribution;
(b)- profile marked in (a).

This phase distribution was combined with the simulated magnitude distribution (see Fig. 5.1(c)) to create a complex image function of the considered object. The fast Fourier transform of this function was then calculated in order to obtain the corresponding simulated complex diffraction pattern function. Figure 5.9 shows in (a) the diffraction pattern intensity distribution and in (b) the indicated profile. Also, in Fig. 5.9(c) and (d) is shown the corresponding simulated magnitude distribution.

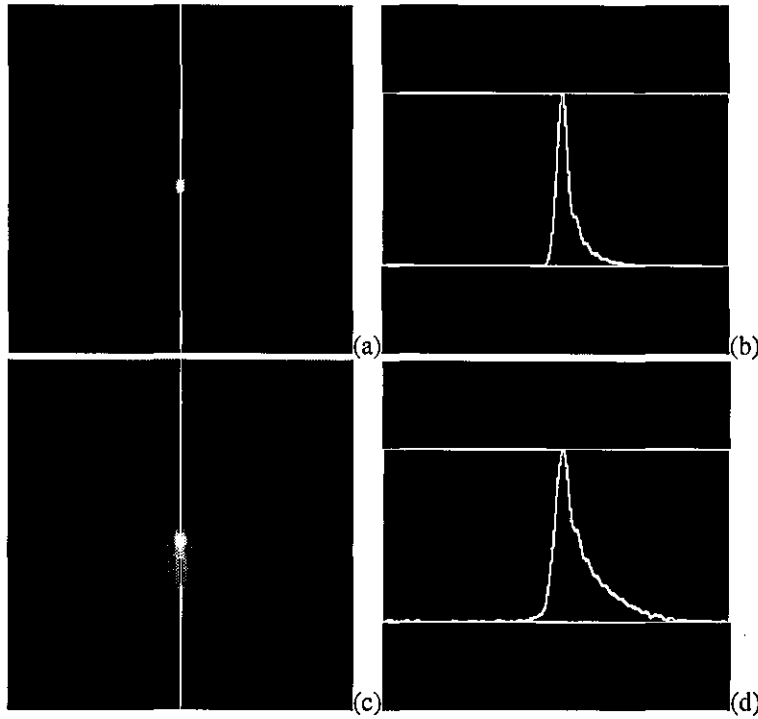


Fig. 5.9 - The diffraction pattern for the exponential disturbance. (a)- intensity distribution; (b)- profile marked in (a); (c)- magnitude distribution; (d)- profile marked in (c).

It should be mentioned that a non-linear phase change in the object or image domain corresponds to a non-linear shift of the diffraction pattern in the Fourier domain. As can be seen in Fig. 5.9, that due to the non-linear phase distribution, the diffraction pattern of the single circular aperture was severely modified and it was also shifted from the central region. In this case only a limited phase change can be measured from the diffraction pattern, and the reasons are exactly the same as mentioned in the previous example. A relation between the phase change and the shift of the diffraction pattern is now more difficult to obtain, when assuming a non-linear phase distribution in the object or image domain.

The objective is now to use these simulated magnitude functions to recover the original phase distribution. The first set of procedures to recover the phase is performed using the magnitude distributions for the image and Fourier domains that have been calculated from the corresponding intensity distributions, respectively shown in Fig. 5.1(a) and Fig. 5.9(a). This set was composed by repeating a 200 iteration process several times feeding the algorithm with the magnitude functions mentioned previously.

All the results obtained in this set were not consistent with each other. Figure 5.10(a) shows one of these recovered wrapped phase distributions when the normalized root mean squared error was 1.99×10^{-1} after 200 iterations.

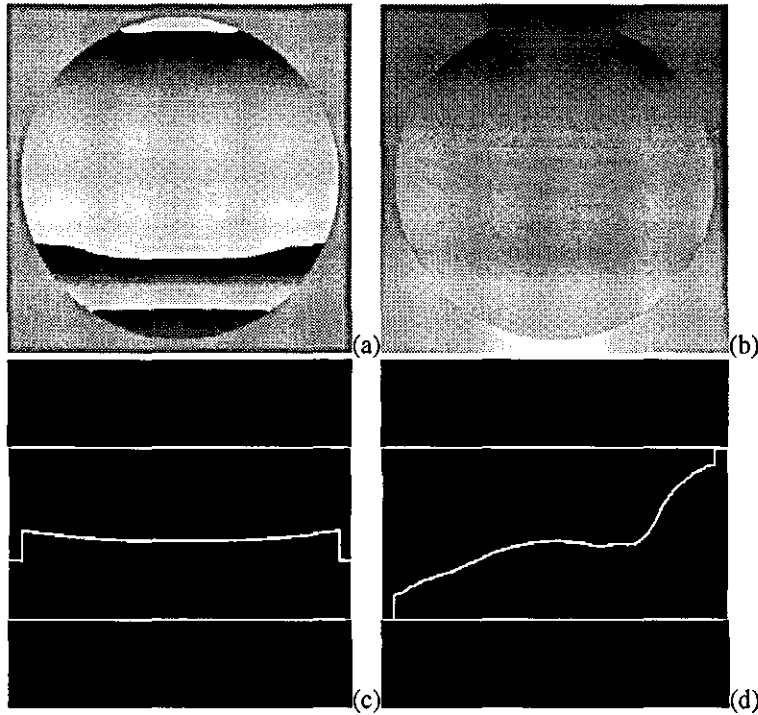


Fig. 5.10 - Recovered phase function for the exponential disturbance. (a)- relative distribution; (b)- absolute distribution; (c)- middle horizontal profile of (b); (d)- middle vertical profile of (b).

After 200 more iterations the error did not change from the previous value which was 1.99×10^{-1} . The corresponding unwrapped phase distribution is shown in Fig. 5.10(b). Also, in Fig. 5.10(c) and (d) are shown respectively the middle horizontal and vertical profiles of the unwrapped phase.

As one can observe from the previous pictures and profiles, the recovered phase distribution does not agree with the original phase distribution and one must consider this set of procedures to retrieve the phase unsuccessful. The reason for the failure has already been mentioned earlier and is concerned with the loss of vital information in the high frequency components of the diffraction pattern function. Also, note that the normalized root mean squared error appears again to be stagnated.

The second set of procedures to recover the phase is performed using the magnitude distributions for the image and Fourier domains that have been simulated, shown respectively in Fig. 5.1(c) and Fig. 5.9(c). This set was also composed by repeating a 200 iteration process several times feeding the algorithm with the magnitude functions previously mentioned.

All the results obtained in this set were now quite consistent with each other. Figure 5.11(a) shows one of these recovered wrapped phase distributions when the normalized root mean squared error was 3.35×10^{-4} after 200 iterations.

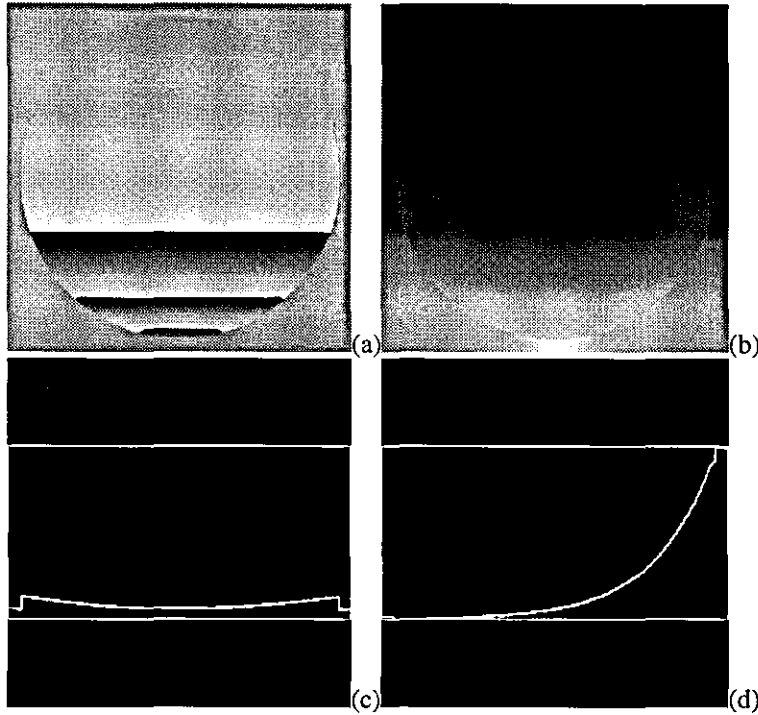


Fig. 5.11 - Recovered phase function for the exponential disturbance. (a)- relative distribution; (b)- absolute distribution; (c)- middle horizontal profile of (b); (d)- middle vertical profile of (b).

After 200 more iterations the error was reduced to 7.91×10^{-6} . The corresponding unwrapped phase distribution is shown in Fig. 5.11(b). Also, in Fig. 5.11(c) and (d) are shown respectively the middle horizontal and vertical profiles of the unwrapped phase.

The results obtained in this set serve to prove once again that the algorithm converges to the correct solution if correct and accurate information is used in the iterative process even for the case of non-linear phase distributions. The normalized root mean squared error shows clearly the convergence of the algorithm. The reproducibility and consistency of the results in all the different processes executed suggests the success of the phase retrieval algorithm to recover the original phase distribution.

Ramp Discontinuity Phase Disturbance

Figure 5.12(a) depicts the simulated phase distribution of a ramp discontinuity viewed through a circular aperture 480 pixels in diameter. The phase change is a linear phase change that increases from the upper-left corner to the lower-right corner of array with a circular discontinuity 240 pixels in diameter at the central region. The phase change along the diagonal corresponds to a change from -3π to $+3\pi$, 6π rad being the total simulated phase change over all the aperture, and π the phase discontinuity jump at the border of the internal circle, as shown by the profile in Fig. 5.12(b).

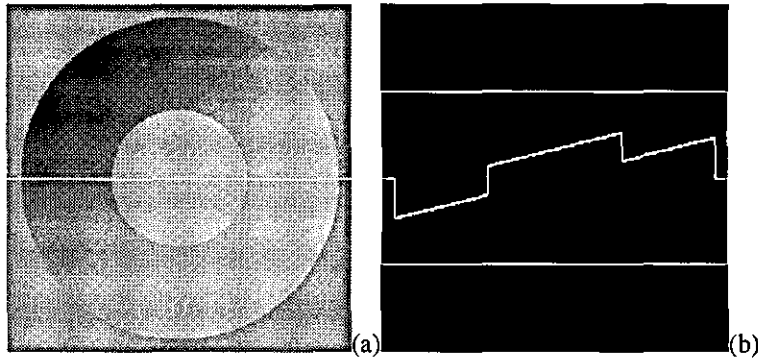


Fig. 5.12 - The simulated phase change for the ramp discontinuity. (a)- phase distribution;
(b)- profile marked in (a).

This phase distribution was combined with the simulated magnitude distribution (see Fig. 5.1(c)) to create a complex image function of the considered object. The fast Fourier transform of this function was then calculated in order to obtain the corresponding simulated complex diffraction pattern function. Figure 5.13 shows in (a) the diffraction pattern intensity distribution and in (b) the profile. Also, in Fig. 5.13(c) and (d) is shown the corresponding simulated magnitude distribution.

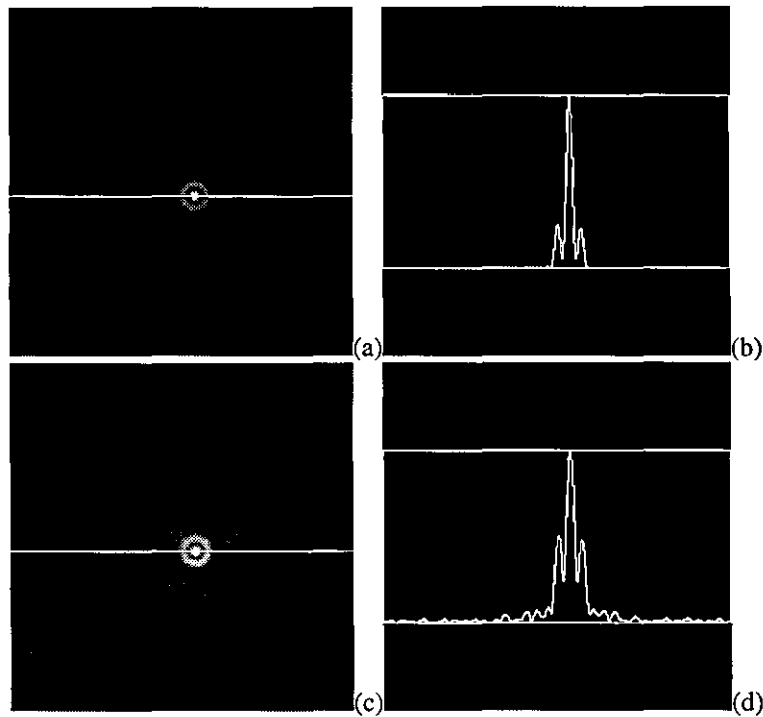


Fig. 5.13 - The diffraction pattern for the ramp discontinuity. (a)- intensity distribution;
(b)- profile marked in (a); (c)- magnitude distribution; (d)- profile marked in (c).

The objective here is to use these simulated magnitude functions to recover the original phase distribution. The first set of procedures to recover the phase is performed using the magnitude distributions for the image and Fourier domains that have been calculated

from the corresponding intensity distributions, respectively shown in Fig. 5.1(a) and Fig. 5.13(a). This set was composed by repeating a 200 iteration process several times feeding the algorithm with the magnitude functions as before.

Again, all the results obtained in this set were not consistent with each other. Figure 5.14(a) shows one of these recovered wrapped phase distributions when the normalized root mean squared error was 2.96×10^{-1} after 200 iterations.

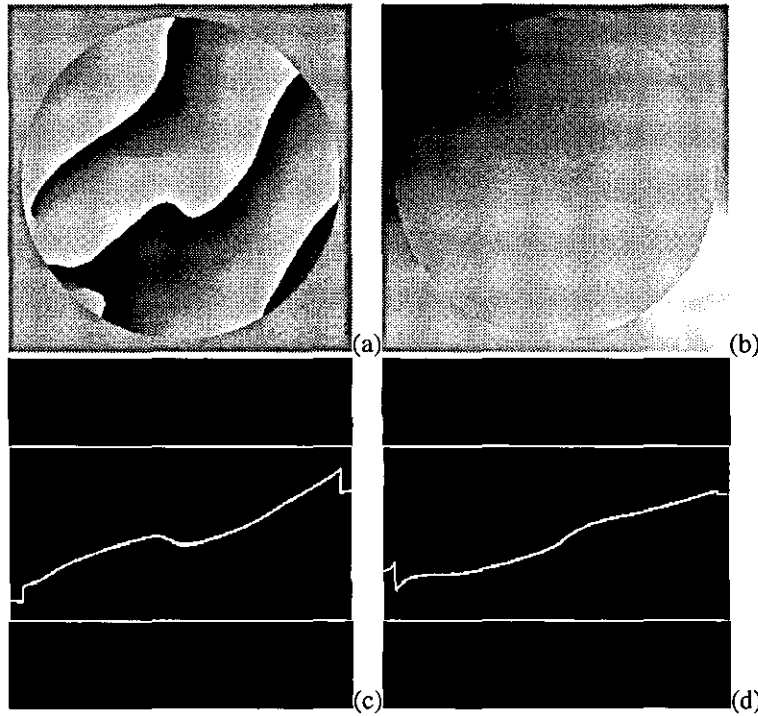


Fig. 5.14 - Recovered phase function for the ramp discontinuity. (a)- relative distribution; (b)- absolute distribution; (c)- middle horizontal profile of (b); (d)- middle vertical profile of (b).

After 200 more iterations the error did not change from the previous value which was 2.96×10^{-1} . The corresponding unwrapped phase distribution is shown in Fig. 5.14(b) and also in Fig. 5.14(c) and (d) are shown respectively the middle horizontal and vertical profiles of the unwrapped phase.

One can observe from the previous results, that the recovered phase distribution does not agree with the original phase distribution and again, this set of procedures to retrieve the phase has been unsuccessful. The reason of the failure is the same as that already mentioned. Also, the normalized root mean squared error appears to be stagnated suggesting that the convergence of the algorithm has stopped in the same way as before.

The second set of procedures to recover the phase is performed using the magnitude distributions for the image and Fourier domains that have been simulated, respectively

shown in Fig. 5.1(c) and Fig. 5.13(c). This set was also produced by repeating a 200 iteration process several times feeding the algorithm with the magnitude functions previously mentioned.

All the results obtained in this set were now quite consistent with each other. Figure 5.15(a) shows one of these recovered wrapped phase distributions when the normalized root mean squared error was 6.37×10^{-3} after 200 iterations.

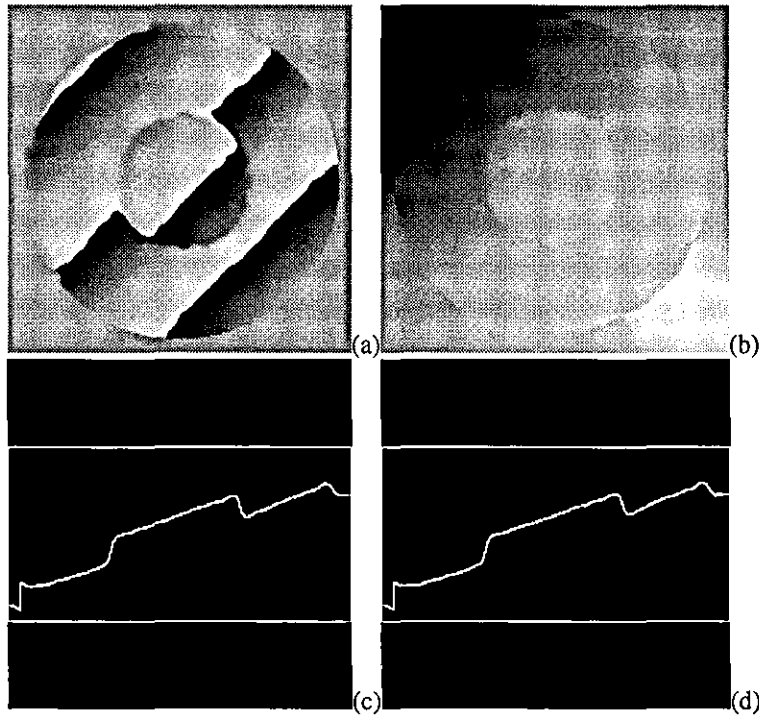


Fig. 5.15 - Recovered phase function for the ramp discontinuity. (a)- relative distribution; (b)- absolute distribution; (c)- middle horizontal profile of (b); (d)- middle vertical profile of (b).

After 200 more iterations the error has been reduced to 2.44×10^{-6} . The corresponding unwrapped phase distribution is shown in Fig. 5.15(b) and also in Fig. 5.15(c) and (d) are shown respectively the middle horizontal and vertical profiles of the unwrapped phase.

The results obtained in this set serve to prove that the algorithm converges to the correct solution if the correct data is used in the iterative process even when discontinuities are present in the phase distribution. It must be mentioned in this particular case that, like in interferogram fringe analysis, phase discontinuities which are multiple of 2π can not be resolved because, the phase is wrapped by the computer into the range $[-\pi, +\pi]$ and consequently the phase unwrapping will produce an incorrect phase distribution. This is in fact another limitation in the phase retrieval process that can not be avoided but, mathematically the wrapped phase is correct and represents the exact phase distribution.

The normalized root mean squared error shows clearly the convergence of the algorithm. The reproducibility and consistency of the results in all the different processes executed suggests the success of the phase retrieval algorithm to recover the original phase distribution.

The above computer simulations proved that the iterative algorithm performs better when true magnitude information from the image and Fourier domains is used instead of the magnitude information calculated from the corresponding intensity records. This performance can be checked any time by analysing the normalized root mean square error given by the algorithm for each iteration. As an example, Fig. 5.16 shows the plot of the normalized rms error versus the number of iterations for a typical phase retrieval process. It should be mentioned that the convergence of the iteration process varies for different processes and it depends the initial phase estimate. For all the simulations that have been carried out during the period of research, it was observed that, in general, the convergence of the algorithm is reasonably fast in the beginning (first 10 to 30 iterations) and it is much slower for the later iterations, in this case extended to 700 iterations.

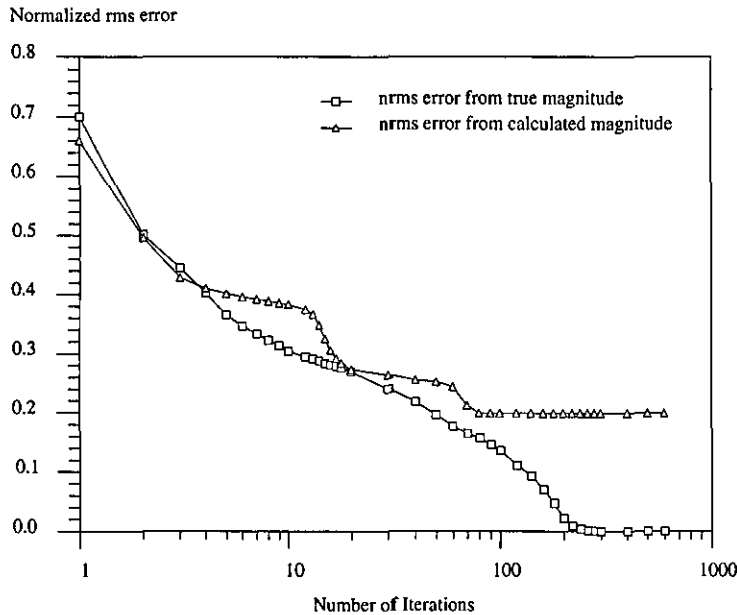


Fig. 5.16 - Normalized root mean square (nrms) error as a function of the number of iterations.

The normalized rms error approaches zero when true magnitude information is considered and it approaches a finite limit greater than zero when magnitude information calculated from intensity is used. This means that the recovered phase distribution approaches a limiting function that is not necessarily the correct solution. Since the true magnitude information will never be available in practical experimentation, this result suggests that care must be taken to measure and record these

quantities and special attention should be given to the Fourier domain, where vital information for the phase reconstruction process can easily be lost due to the low intensity levels usually involved.

5.3.3 Conclusions

The computer simulations have proved that the iterative algorithm for phase retrieval can be used successfully to recover the phase distribution associated with an object knowing the magnitude distributions of its image and diffraction pattern. The convergence of the iteration process has been carefully analysed and the conditions for a phase retrieval success have also been discussed in terms of accuracy of the measurements of the required magnitude information from the image and Fourier domains of concern.

Different object disturbances have been simulated to show the performance of the retrieval process when the phase distributions were respectively linear and non-linear. Also the problem of discontinuities in the phase distribution have been considered. The results obtained are of great value and this method can be used on the reconstruction and characterization of an object disturbance.

The ability to recover by simulations phase distributions, which contain most of the important information about an object, is an invitation to extend and apply such procedures to real physical situations. An important point that should be highlighted concerns the limitations that have been found from the computer simulations carried out. It has been found that the method can not be used under the present conditions to recover phase changes larger than approximately $12 \times 2\pi$ rad or in other words, when the phase gradient over the whole system aperture exceeds $12 \times 2\pi$ rad. A scale compromise between the image and Fourier domains has to be made and in principle this method can be applied for any kind and amount of phase change. The use of large arrays is the solution to increase the amount of phase change that can be recovered, but the time of processing also increases considerably perhaps making the method impractical.

Independent of the limitations mentioned above, this method, as it has been described in the previous sections, can be used to recover small phase changes and the resolution that can be achieved to detect phase changes is $2\pi/20$ rad, giving to the method a very high sensitivity.

5.4 Experimentation

The research that has been carried out concerns the phase retrieval and therefore the reconstruction of complex wave-fronts for simple real object disturbances. It involves the application of the same algorithm, that have been used on the simulations presented in the previous section, where the input information is now taken from real measurements of intensity in an experimental apparatus. In order to extend the simulated results to real experimental results an experimental arrangement has been built, which is fully described in the next section. The experimental results obtained with the apparatus for a set of simple phase objects are presented and discussed.

5.4.1 Experimental Arrangement

Figure 5.17 shows schematically the experimental apparatus that enables the measuring and recording of wave-front intensities, in the image and Fourier domains, created by an object disturbance.

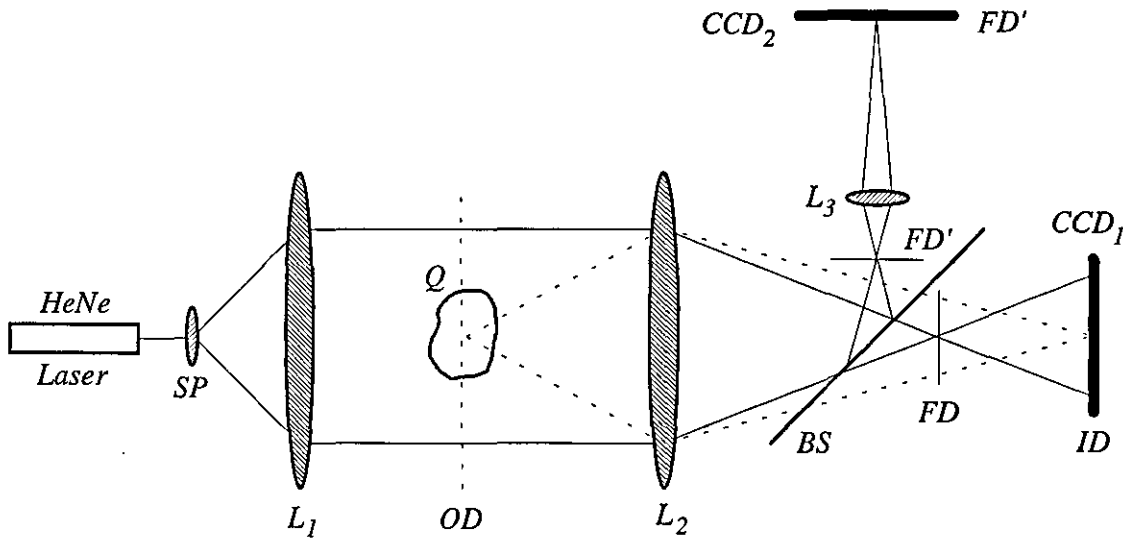


Fig. 5.17 - Experimental arrangement for recording the wave-front intensity distributions in the image and Fourier domains produced by an object disturbance.

It uses a HeNe laser with wavelength $\lambda = 632.8$ nm, a spatial filter *SP* and the collimating lens *L*₁ to produce a plane parallel beam of light. This monochromatic beam is used to illuminate the object disturbance *Q* placed in the object domain *OD*. The object disturbance will introduce some phase changes in the wave-front that travels through it due to changes in the refractive index or equivalently changes in the optical path length in the object field.

All the light passing through and deviated by the object is collected by the Fourier lens L_2 , which has a large aperture, and the light intensity changes are imaged in the image domain ID .

For acquisition and recording of the image intensity distribution a solid state sensor CCD_1 (Charge Coupled Device) has been placed in the image domain. The Fourier domain FD created by the lens L_2 contains the Fraunhofer diffraction pattern of the object disturbance. This plane is mirrored to FD' through a beam-splitter cube BS and the light intensity information on it is imaged in another solid state sensor CCD_2 using a microscope objective lens L_3 . Both the CCD sensors are connected to a computer workstation through a frame grabber to perform multi-channel monochrome image acquisition, as described in chapter two Section 2.4. Also, the application program, specifically developed to support this research, is used to acquire and record images on the computer as well as to manipulate and process all the recorded information.

5.4.2 Set-up, Scaling and Sampling

The beam from the HeNe laser is focused using a 10x microscope objective and spatially filtered with a 25 μm pinhole. This pinhole is placed in the focal plane of the microscope objective and the result is an expanded clear spatially filtered laser beam. The expanded beam is then collimated by the collimating lens L_1 , with focal length $f_1 = 120$ mm, to produce a circular plane parallel beam of monochromatic light with $D = 16$ mm in diameter.

The parallel beam of light is used to illuminate an object disturbance and then all the light passing through and deviated by the object is collected by the large aperture Fourier lens L_2 , with focal length $f_2 = 120$ mm, to produce an image of the object in the image plane, as indicated in Fig. 5.18. The object disturbance is placed in the object plane, also called the object domain OD , located at S_o from the Fourier lens. The image plane, also called the image domain ID , is located at S_i from the Fourier lens where the sensor CCD_1 is located to acquire and record the light intensity distribution in this domain. Since the dimensions of the active area of the CCD array sensors are $6.4 \times 4.9 = (752 \text{ pixels each } 8.5 \mu\text{m}) \times (582 \text{ pixels each } 8.5 \mu\text{m})$ mm, the image of the circular object aperture D was de-magnified to fill the area of the sensor. One can obtain a 4x de-magnification by making $M = 0.25$ in the geometrical optics relation [4,5]

$$M = \frac{S_i}{S_o}, \quad (5.4-1)$$

where M is the image magnification, and S_i , S_o are respectively the distance between the image plane and lens L_2 and the distance between the lens L_2 and the object plane. In this way the aperture $D=16$ mm will be imaged on the sensor CCD_1 as a 4 mm circular disc.

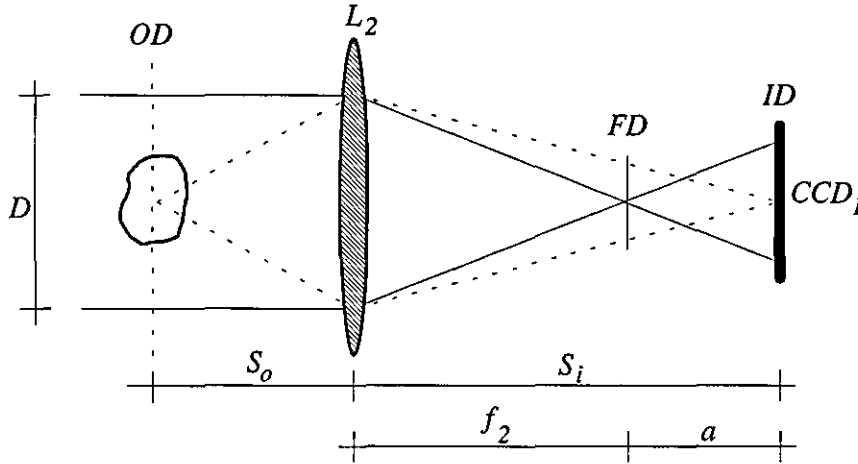


Fig. 5.18 - Experimental arrangement. Image domain set-up.

The distances S_i and S_o are also related to the focal length f_2 of the Fourier lens L_2 by the geometrical optics relation [4,5]

$$\frac{1}{S_i} + \frac{1}{S_o} = \frac{1}{f_2}. \quad (5.4-2)$$

From Eqs. (5.4-2) and (5.4-1) one can calculate the distances S_i and S_o in terms of the magnification M and the focal length f_2 . For the above conditions the results are $S_o = 600$ mm and $S_i = 150$ mm, so that the distance $a = 30$ mm. It should be noted that the distances indicated in the diagram in Fig. 5.18 are not to scale.

The Fourier plane, also called the Fourier domain FD , is mirrored to FD' using a beam-splitter cube BS , with 30-30% reflectance/transmittance. As well-known from the Fourier optics and Fourier transform theory [3], the scale in the Fourier domain is inversely proportional to the scale in the object or image domain. This means that the Fourier domain information must be recorded with a magnification inversely proportional to the magnification used in the image domain. In other words, if detailed intensity information in the Fourier domain is required the corresponding intensity information in the image domain should be de-magnified accordingly or if detailed

intensity information in the image domain is required the intensity information in the Fourier domain should be de-magnified proportionally.

Since the information in the image domain has been de-magnified 4x ($M = 0.25$) then the information in the Fourier domain must be magnified 4x ($M = 4$). It has been found experimentally that a magnification of 4x for the information in the Fourier domain results in the recording of a diffraction pattern distribution with no detail because it is confined to a small area of about 2×2 pixels on the CCD array. For this reason a compromise needs to be made when these images are acquired from the experiment. In this case in particular, one is looking for detailed intensity information in the Fourier domain and therefore a 20x magnification was selected due to practical limitations on the experiment and a 20x microscope objective was used in the apparatus as indicated by L_3 in Figs. 5.17 and 5.19.

The first practical limitation is related to the object aperture, which has been fixed at $D = 16$ mm. If more than 20x magnification is used for the Fourier domain, the diffraction pattern is too magnified and extends beyond the available area on the sensor CCD_2 , and therefore important information concerning the high frequency components on the recorded diffraction pattern is missing. On the other hand, if less than 20x is used the fine detail on the low order frequency components is completely lost and small changes in the shape of the diffraction pattern can not be detected. The 20x magnification was chosen as a compromise because of the need to not lose parts of the diffraction pattern due to image shifts out of the available area on the CCD sensor. Since the shifts of the diffraction pattern are related to phase changes introduced by the object disturbance on the wave-front, as mentioned in chapter three and also in the previous section, with this system the measurement of large wave-front phase changes is limited by the size of the sensor.

The second limitation concerns the inverse scale between the image and Fourier domains (see Sections 3.3 and 3.4 in chapter three). If a large magnification is selected for the Fourier domain the corresponding inverse magnification or de-magnification must be achieved for the image domain, in order to obtain on the computer two scaled functions which are related to each other by the Fourier transformation. Being in scale, these two functions can be transformed and manipulated directly without additional scaling problems. The assumptions made above about the system give a Fourier domain magnification of 20x and an image domain de-magnification of 4x, therefore, it is only necessary to have an extra de-magnification of 5x of the information in the image domain to obtain the required scale agreement. For practical reasons, this 5x de-

magnification is performed on the computer using the appropriate algorithm for de-magnification included on the developed application program. The input data to the algorithm consists of the recorded information taken from the image domain in the experiment and the (de-)magnification factor.

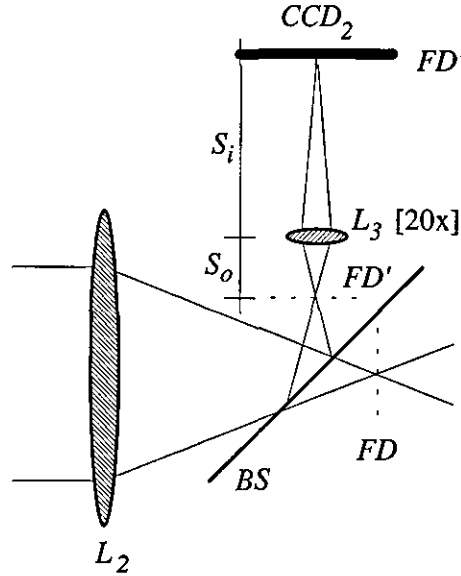


Fig. 5.19 - Experimental arrangement. Fourier domain set-up.

The 20x microscope objective is mounted on a 160 mm tube, that is the distance $S_i = 160$ mm indicated in Fig. 5.19. Using the same relations expressed in Eqs. (5.4-1) and (5.4-2), and knowing that $M = 20$ and the value of S_i , one can obtain the value $S_o = 8$ mm. Note that the microscope objective lens L_3 has a focal length $f_3 = 7.5$ mm. In order to verify in practice the Fourier domain magnification a simple experiment has been carried out with the above apparatus. A coarse grating with $40 \mu\text{m}$ strip width (Coarse Grating 2) has been placed in the Fourier plane identified by FD' in Fig. 5.19. White light has been used to produce on the sensor CCD_2 an image of the grating. The obtained image has been recorded on the computer and measurements of the grating spacing have been made. Each strip was found to occupy 94 pixels. This gives a width $w = 94 \times 8.5 \times 10^{-6} = 799 \mu\text{m}$, using the value $8.5 \times 8.5 \mu\text{m}$ for the pixel size, which is in fact $19.975 \approx 20x$ bigger.

As mentioned earlier, the CCD sensors are connected to a computer through a frame grabber for acquisition and recording purposes. Another important aspect concerning these two sensors is related to their physical positioning and alignment on the experimental apparatus. Both the sensors are mounted in XYZ holders making possible accurate movements in all directions in space. The accurate position of the sensors involves, firstly, manual XYZ positioning, and secondly, computer XY movements

using the facility developed for independent pan and scroll movements of the acquired images on the screen. The precision achieved is one pixel in both the vertical and horizontal directions and is totally related to the pixel size on the CCD arrays, which is $8.5 \times 8.5 \mu\text{m}$ for this particular case (see Section 2.4 in chapter two for more details).

A relation between the computer Fourier transformation and the optical Fourier transformation is necessary at this stage to understand completely the scaling problems. By the Fraunhofer diffraction theory [6], the distance d between successive minima and between the central maximum and the first minimum in the diffraction pattern at the focus of a lens of focal length f by a circular aperture with diameter D uniformly illuminated by monochromatic light λ , as shown in Fig. 5.20, can be expressed by the following equation

$$d = 1.22 \frac{\lambda f}{D}. \quad (5.4-3)$$

This equation can be generalized for the discrete case, where the distances are measured in terms of number of pixels and the product λf is substituted by the total number of pixels of the considered sample.

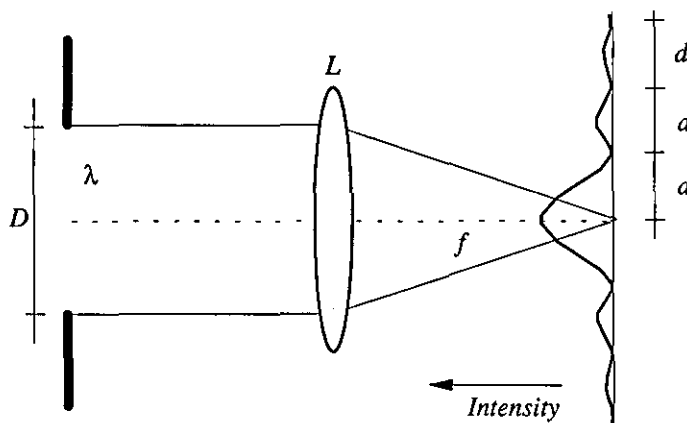


Fig. 5.20 - Distance between successive minima on the diffraction pattern of a circular aperture.

For example, for a sample of $N = 512$ pixels the corresponding distance r (in pixels) between successive minima and between the central maximum and the first minimum in the diffraction pattern of a circular aperture with n pixels in diameter can be expressed by

$$r = 1.22 \frac{N}{n}. \quad (5.4-4)$$

The equivalent distance in meters can be easily obtained by multiplying the number of pixels r by the physical size of each pixel.

In order to find the relation between the computer Fourier transformation and the optical Fourier transformation, consider a 512×512 pixel image where a small circle, 48 pixels in diameter, with uniform intensity distribution has been placed in the centre of the picture as shown in Fig. 5.21(a). The discrete Fourier transform of such an image has been calculated on the computer using the developed fast Fourier algorithm to obtain the Fraunhofer diffraction pattern of the aperture whose profile is shown in Fig. 5.21(b).

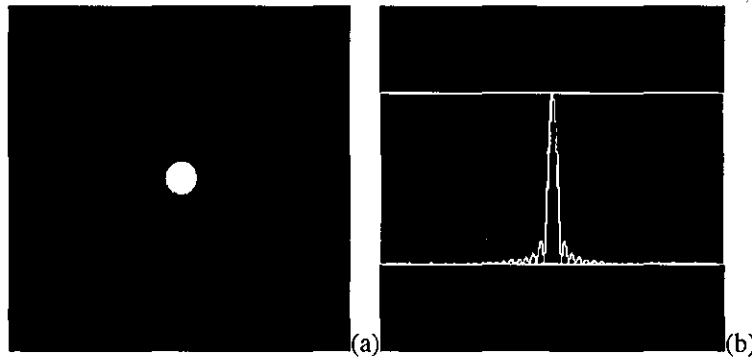


Fig. 5.21 - The discrete Fourier transform of a circular aperture. (a)- circular aperture distribution; (b)- profile of the calculated diffraction pattern.

Using this profile one can easily measure on the computer the distance between the central maximum and the first minimum in terms of pixels. The measured distance was $r = 13$ pixels which agrees with the calculated distance if Eq. (5.4-4) is used with $N = 512$ and $n = 48$.

Suppose now, that the image of the circular aperture has been acquired and recorded on the computer using the experimental apparatus described above. For this particular case and knowing that the size of each pixel on the CCD sensors is $8.5 \mu\text{m}$, one can obtain the physical dimensions of the circular aperture which is $D = 48 \times 8.5 \times 10^{-6} \approx 0.4 \text{ mm}$. Using Eq. (5.4-3) with $\lambda = 632.8 \text{ nm}$ and $f = 120 \text{ mm}$ the corresponding theoretical distance is $d_{0.4} = 231.6 \mu\text{m}$. This does not agree with calculated distance using the value of r , which is $d_{0.4} = r \times 8.5 \times 10^{-6} \approx 110.6 \mu\text{m}$.

From the above result one can say that the computer Fourier transformation is not calibrated with the theoretical and experimental optical Fourier transformation. Since the wavelength λ is a fixed parameter in the experiment, the calibration problem is to find out what is the focal length f of the *lens* used by the computer when it calculates a Fourier transformation. Using the value of d , calculated above using the number of

pixels r multiplied by the pixel size, and the value of λ on Eq. (5.4-3), one can obtain the focal length for the *computer transformer lens* which is $f \approx 60$ mm. The above procedures have been repeated for different diameters of the circular aperture and the results obtained were all consistent and they were in perfect agreement with the sampling theorem, which has been discussed in chapter three Section 3.4.

In this way it has been found that the computer with arrays of 512×512 pixels elements, which corresponds to a square array with edge 4.352 mm length, and using the wavelength $\lambda = 632.8$ nm, performs the same transformation as a lens with focal length $f = 60$ mm. This value is half of the focal length of the lens L_2 on the experimental apparatus. Using Fourier optics theory [6,7], one can state for this particular case that, from the computer's point of view the intensity information in the Fourier domain is magnified 40x on the sensor CCD_2 and the intensity information in the image domain, as required for practical reasons, is de-magnified 4x on the sensor CCD_1 . In order to obtain the corresponding inverse scale between the intensity information in the Fourier domain and the intensity information in the image domain and therefore the desired calibration of this system, the picture taken from CCD_1 must be de-magnified, on the computer, 10x more to obtain a compatible scale. This can be verified if Eq. (5.4-3) is evaluated for an aperture diameter $D = 16$ mm, $\lambda = 632.8$ nm, and $f = 120$ mm. The resulting distance is $d_{16} = 5.79 \mu\text{m}$, which is in fact 40x smaller than the value obtained before, which is $d_{0.4} = d_{16} \times 40 = 5.79 \times 10^{-6} \times 40 = 231.6 \mu\text{m}$, when the diameter of the circular aperture was set to 48 pixels equivalent to 0.4 mm.

It should be noted, in this particular case, that the real dimensions of the diffraction pattern which are indeed very small if compared with the size of an individual pixel, make clear the necessity for magnified Fourier diffraction pattern information with a compromise for the de-magnification of the image information.

The acquisition and recording of intensity information taken from the image and Fourier domains, in the experiment, have shown two other problems. These were related to the background noise generated by dark-currents which are always present in solid state CCD sensors, and to the pixel saturation and blooming due to the high intensity of the laser light used. The first of these problems, the background noise, is easily reduced by proper monitoring and controlling of the sample and clamp levels of the analogue to digital converters in the frame grabber. The developed program has the capability to monitor and control in real time these key parameters during the digitization process in order to reduce to a minimum undesired effects. Also, a special algorithm based on low pass filtering can be applied to the recorded data to smooth noise effects and

consequently to obtain clear data pictures. To overcome the second problem of saturation and pixel blooming, the shutter speed on the CCD sensors must be adjusted for individual cases and also manual gain control must be used and adjusted instead of the automatic gain control. A rotating switch can be set on the cameras in one of eight positions to obtain different shutter speeds. In particular, the shutter speed on the sensor CCD_1 (image domain) was set to $1/125$ sec, and on the sensor CCD_2 (Fourier domain) was set to $1/4,000$ sec.

After setting up the experimental apparatus, one can acquire and record intensity information on the computer, from the image and Fourier domains, for later processing and analysis. As a conclusive example, Fig. 5.22 shows, respectively in (a) and (b), the intensity information taken from the Fourier domain and the marked horizontal profile, when no object disturbance is placed in the object domain.

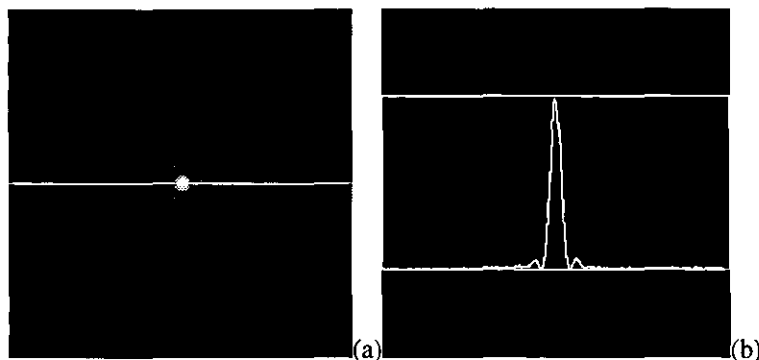


Fig. 5.22 - The experimental diffraction pattern of the real circular aperture.

(a)- intensity distribution; (b)- profile marked in (a).

In order to verify experimentally the theoretical value for the distance calculated by Eq. (5.4-3), the profile shown in Fig. 5.22(b) has been used to measure the number of pixels r between the central maximum and the first minimum, which was $r = 22$ pixels. Using the same procedure, the correct distance d_{16} can be obtained by multiplying the value of r by the pixel size divided by 20 (magnification factor) giving the value $d_{16} = 9.35 \mu\text{m}$. It has been shown before that the theoretical value was $5.79 \mu\text{m}$ and not $9.35 \mu\text{m}$. Why this difference now? The answer to this question is that the difference is totally related to the use of an inappropriate equation, Eq. (5.4-3), to calculate the theoretical value. The referred equation is only valid when a circular aperture is uniformly illuminated. In practice, the real circular aperture has a non-uniform intensity distribution as is clearly visible on the picture shown in Fig. 5.23(a) and on the corresponding horizontal profile displayed in Fig. 5.23(b). This picture was taken simultaneously from the image domain with the associated diffraction pattern in the Fourier domain which is shown in Fig. 5.22(a).

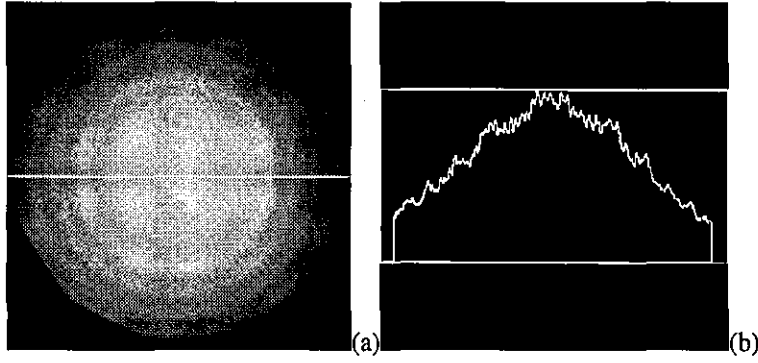


Fig. 5.23 - The experimental image of the real circular aperture.

(a)- intensity distribution; (b)- profile marked in (a).

From Fig. 5.23, one can say that the intensity distribution in the object or image domain is a truncated Gaussian distribution of amplitudes as shown by the model depicted in Fig. 5.24(a) and (b) and the corresponding profiles superimposed in (c).

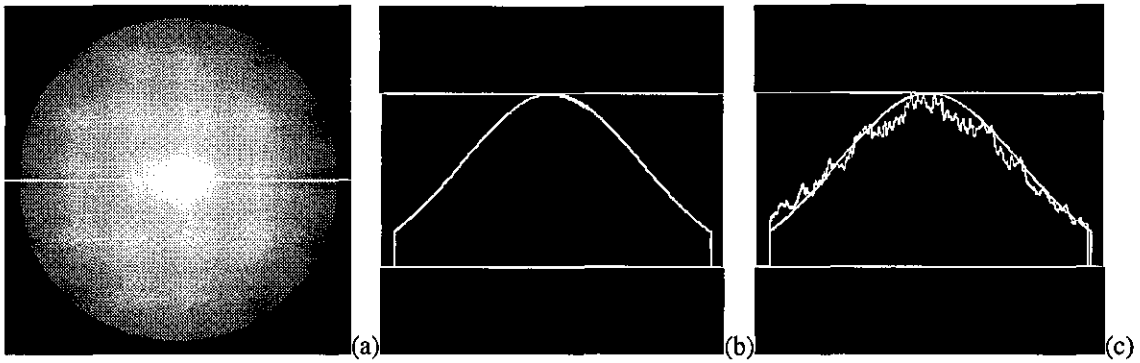


Fig. 5.24 - The modelled image of the circular aperture. (a)- intensity distribution; (b)- profile marked in (a); (c)- superposition of the experimental and modelled profiles.

For non uniform distributions of intensity the problem must be treated carefully. However, for the particular case of Gaussian distributions it can be shown [6], that the distance d between the central maximum and the first minimum can be expressed by the following equation

$$d = 2 \frac{\lambda f}{D}, \quad (5.4-5)$$

and consequently the Eq. (5.4-4) should be rewritten for the case of Gaussian distributions as

$$r = 2 \frac{N}{n}. \quad (5.4-6)$$

Using the above equations for the same conditions, one can obtain $d_{16} = 9.49 \text{ } \mu\text{m}$ and $r = 21.33 \approx 22$ pixels as it should be.

At this stage it should be mentioned the reason why in the previous section, which is dedicated to computer simulations, truncated Gaussian distributions have always been used for the intensity in the image domain. The reason, as mentioned in Section 5.3, is simply to be as close as possible to the reality of the experimental conditions on the apparatus used throughout this section.

5.4.3 Intensity and Quasi Magnitude Measurements

As pointed out in Section 5.3, the phase retrieval algorithm gives the correct phase solutions without any ambiguity if the input data on the algorithm for both the image and Fourier domains corresponds to the exact magnitude information in such domains. This means that the algorithm, as demonstrated, does not give correct solutions for the phase problem when the input magnitude data has been calculated from simulated intensity distributions. Approximate and ambiguous solutions with convergence stagnation have been obtained in the phase retrieval process due to loss of information during the calculation of the input data to the iterative algorithm. It is clear from these simulated results that care must be taken when the corresponding magnitude distributions have to be calculated in order to feed adequately the iterative phase retrieval algorithm.

Since the magnitude distribution $|g(u, v)|$ of a complex quantity $g(u, v)$ is the square root of the measured intensity distribution $I_g(u, v)$, defined as $|g(u, v)| = \sqrt{I_g(u, v)}$, one can not obtain exact magnitude distributions from the corresponding intensity distributions. There are several reasons for this fact. The first reason concerns the limited range in the digitization process which is 256 integer values to quantify information where strong and weak amplitude variations exist simultaneously. Normally these weak amplitude variations are converted to zero values and only the strong amplitude variations are reasonably converted within the available range. In fact, the information in the Fourier domain is an example of such weak and strong amplitude variations, and problems in accurate measurements have to be faced. When this inaccurate intensity information, which is obtained from the CCD sensors, is used to calculate the corresponding magnitude information the resulting magnitude distribution is also inaccurate. The second reason concerns the truncation and rounding operations performed on the digitization process, that is, when an analogue quantity is discretized the intermediate

values are truncated and rounded to integer values incurring a loss of information. Also, loss of information is created by precision errors that are introduced in the calculation of the magnitude data. They are related to the square rooting operation where an integer value from the intensity data is used to calculate a real value to represent the magnitude data. Bearing in mind these two reasons, it will be very difficult to obtain in practice magnitude distributions with the same or equivalent accuracy as the magnitude distributions used in the previous simulations.

By coincidence, the solid state CCD sensors used on the experimental apparatus have the capability to perform electrically an operation called the *gamma correction*, where γ is the electrical analogue of the response of a photographic emulsion [8], that is the ratio of the output voltage on the CCD sensors is proportional to the *logarithm* of the light intensity, as discussed in chapter two. A switch, that is available in each of the CCD sensors, can be set to $\gamma=1$ and a linear transformation between light intensity and output voltage is performed or it can be set to $\gamma=0.45$ and therefore a non-linear transformation performed.

A simple experiment has been carried out using the actual set up to demonstrate this capability. A rotating polarizer has been placed across the collimated laser beam on the apparatus to change the intensity of the transmitted light which can be measured and recorded in the image domain by the sensor CCD_1 . The transmitted light intensity I_t is proportional to the square of the cosine of the incident light intensity I_o and it can be represented by the relation

$$I_t = I_o \cos^2 \phi, \quad (5.4-7)$$

where ϕ is the angle between the position of the polarizer axis and the polarization axis on the laser beam. Since the position of the polarization axis on the laser beam is known, the polarizer was rotated from 0° to 90° in steps of 2° and a set of measurements has been performed in the image domain for each of the γ values. The measurements showed excellent reproducibility in that the changes in the measured intensities did not differ by more than 1% from each other. It has been found that no changes in the intensity for angles between 0° and 18° and also for angles greater than 46° in the two sets of measurements that have been performed. The results obtained for the normalized averaged measurements of the output voltage on the sensor versus the square of the cosine of the angle, changing between 18° and 46° , for these two sets of data, one for each γ value, are displayed in Fig. 5.25(a).

From these curves it is clear that a linear transformation is performed by the sensor when $\gamma = 1$ and a non-linear transformation is performed when $\gamma = 0.45$. Figure 5.25(b) shows the relation between the normalized averaged output voltage for $\gamma = 0.45$ on the CCD sensor and the simulated normalized output voltage for an estimated value $\gamma = 0.54$. The estimated value $\gamma = 0.54$ has been obtained from a computer program for curve fitting using the normalized output voltage data when the CCD was set to perform gamma correction of 0.45. The above result shows that an error of about 17% exists in the gamma value indicated in the technical specifications of the CCD sensor.

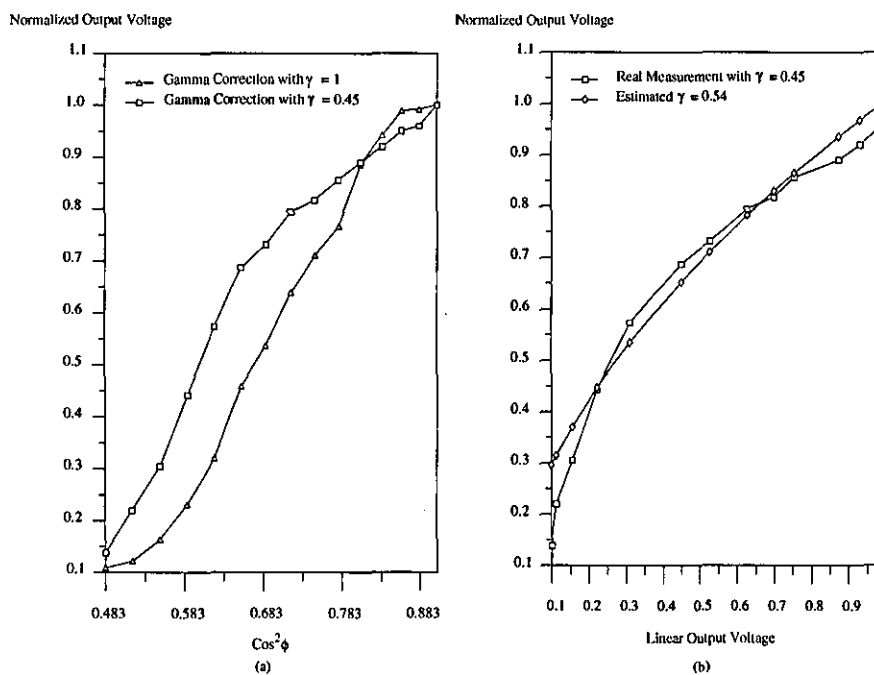


Fig. 5.25 - Averaged output voltage of transmitted intensity. (a)- for gamma equal to 1 and 0.45;
(b)- comparison between real and estimated data.

This non-linear transformation gives an increased output voltage on the CCD sensors for low intensities of light based on the measured intensity raised to the power of γ . Thus, the measuring and recording of low intensity information will be possible and more accurate measurements can be made. Since the magnitude is the square root of the measured intensity or in other words, the magnitude is the measured intensity raised to the power of 0.5, one can state that, if the CCD sensor is set to perform the gamma correction for $\gamma = 0.45$, or more precisely $\gamma = 0.54$, the information available on the output of this sensor is the *quasi* magnitude information, which can mathematically be expressed by the following relations

$$\begin{aligned} |g(u, v)| &= \sqrt{I_g(u, v)} = I_g(u, v)^{0.5} \\ |g(u, v)| &\approx I_g(u, v)^\gamma = I_g(u, v)^{0.54} \end{aligned} \quad (5.4-8)$$

and consequently an error of about 7.5% exists between the recorded quasi magnitude data and the true magnitude data, as depicted in Fig. 5.25(b).

In this way, one can use this capability to acquire and record on the computer the quasi magnitude of the diffraction pattern, if the sensor CCD_2 on the experimental apparatus is set to perform gamma correction of $\gamma = 0.45$ ($\gamma = 0.54$). Then by processing in the computer the recorded data, the desired magnitude data can be obtained and consequently more accurate measurements of the diffraction pattern magnitude can be made directly from the CCD sensor. Figure 5.26(a) and (b) shows the recorded quasi magnitude distribution and the corresponding horizontal profile of the diffraction pattern taken from the experimental set up with no disturbance in the object domain.

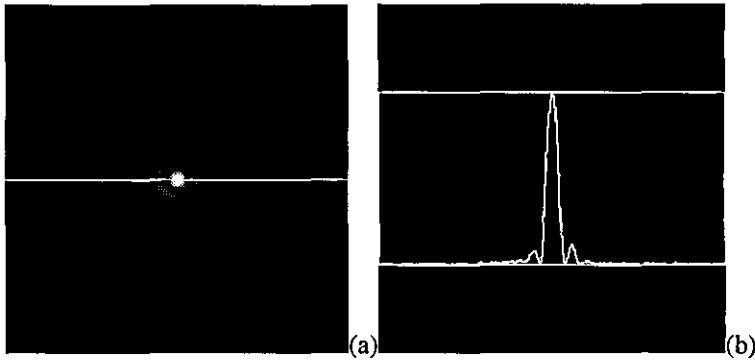


Fig. 5.26 - The experimental diffraction pattern of the real circular aperture.

(a)- quasi magnitude distribution; (b)- profile marked in (a).

Since the accuracy of the measurements in the image domain is not as critical as that of the measurements in the Fourier domain, the sensor CCD_1 is normally set to perform a linear transformation and the normal procedure is taken on the recorded intensity data to obtain the corresponding magnitude information, as shown in Fig. 5.27, in (a) the calculated magnitude from the recorded intensity and in (b) the profile.

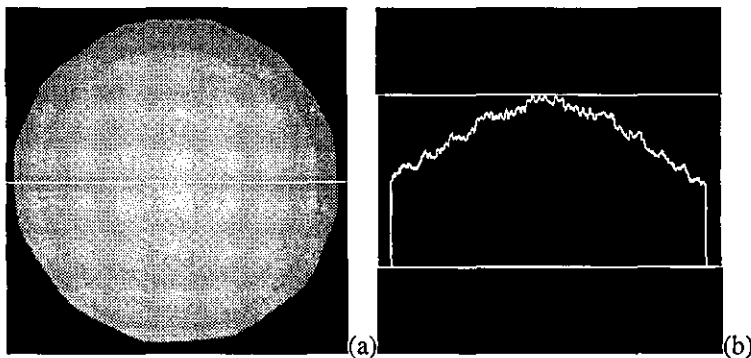


Fig. 5.27 - The experimental image of the real circular aperture.

(a)- magnitude distribution; (b)- profile marked in (a).

The next section is devoted to the presentation and discussion of the experimental results obtained with the described apparatus to retrieve the phase information associated with complex wave-fronts created by a set of real object disturbances.

5.4.4 Experimental Results

The iterative algorithm will be used to retrieve the phase distribution of a real object wave-front. The input data to the iterative algorithm is now calculated from a set of measurements taken from the experimental arrangement. The first measurement concerns the recording of the intensity information in the image domain, the image of the object disturbance, and the second concerns the measurement of the quasi magnitude information in the Fourier domain, the diffraction pattern of the corresponding object.

The acquired data is stored on the computer and it is used to calculate the corresponding magnitude distribution functions, and also the necessary 10x de-magnification of the magnitude distribution in the image plane for scale agreement purposes. Since the relation defined between the magnitudes in the image and Fourier domains by the Parseval's theorem must be satisfied (see chapter four Section 4.4.2), these functions are then modified by amplitude re-scaling to guarantee the conservation of energy in the Fourier transformation process.

Another important relation to keep in mind is the orientation of the co-ordinate systems (x,y) and (u,v) associated with these functions, for the image and Fourier domains respectively. This means that their relative position must be maintained to obtain exact agreement between the considered data. Bearing in mind that the CCD sensors invert and flip the pictures projected on their active sensing area, it should be noted that due to beam splitter *BS* (see Fig. 5.17) the information in the Fourier domain denoted by FD' is flipped with respect the information in FD and it is therefore necessary to perform on the computer the inversion of the Fourier magnitude function. After all these operations, these modified functions are the effective input functions to the algorithm.

The algorithm iteratively estimates the phase distribution function in one of the selected domains. For this case in particular, it will always be the image domain. The reason for this selection is related to the physical interpretation of the image phase distribution, as discussed in chapters one and four. The physical interpretation of the Fourier phase distribution is not always clear or as easy as the interpretation of the image or object phase distribution. The computer for each iteration executes one direct Fourier

transformation and inverse Fourier transformation pair over the input data. This procedure can be thought of in terms of the image formation process as the computer operating on the input data in exactly the same way as the lens L_2 , the Fourier lens, operates in the experiment (see Fig. 5.17). The iterative procedure considered here always starts by generating a random phase function for the Fourier domain which is uniformly distributed in the interval $[-\pi, +\pi]$, as described in chapter four Section 4.4.2. It is possible to start the iteration process by using a guess for the phase distribution instead of the random phase if a sensible or intuitive guess can be made.

Since the algorithm is an iterative procedure the convergence of the algorithm will be monitored through the calculated normalized root mean squared error, as mentioned in chapter four Section 4.4.2. The number of iterations executed will vary for different retrieval processes, but in general no more than 200 iterations are performed to obtain one solution. The timing for each iteration is basically twice the time of a single discrete Fourier transformation, which is about 9 seconds, giving a maximum of 30 minutes for a 200 iteration process.

The following text concerns the presentation and discussion of the experimental results obtained with the above system for a set of phase object disturbances. This set consists of a glass slide, a soap film, a heat flow and finally for no object disturbance but one in which the optical system itself is investigated.

Glass Slide

The aim here is to try and measure the optical inhomogeneities in a mass produced glass microscope slide. Figure 5.28 depicts the recorded data associated with the distribution functions taken from the image and Fourier domains in the experimental arrangement when the glass slide was placed in the object domain OD to produce an optical wave-front disturbance. In particular, the object magnitude function obtained from CCD_1 is shown in Fig. 5.28(a) and the corresponding profile in Fig. 5.28(b). The quasi magnitude function of the diffraction pattern taken from CCD_2 is shown in Fig. 5.28(c) and in (d) its horizontal profile.

A set of procedures to retrieve the phase distribution in the image domain, for this real object, has been executed. This set was generated by repeating a 200 iteration process several times feeding the algorithm with the magnitude functions previously calculated using the steps described in the beginning of this section.

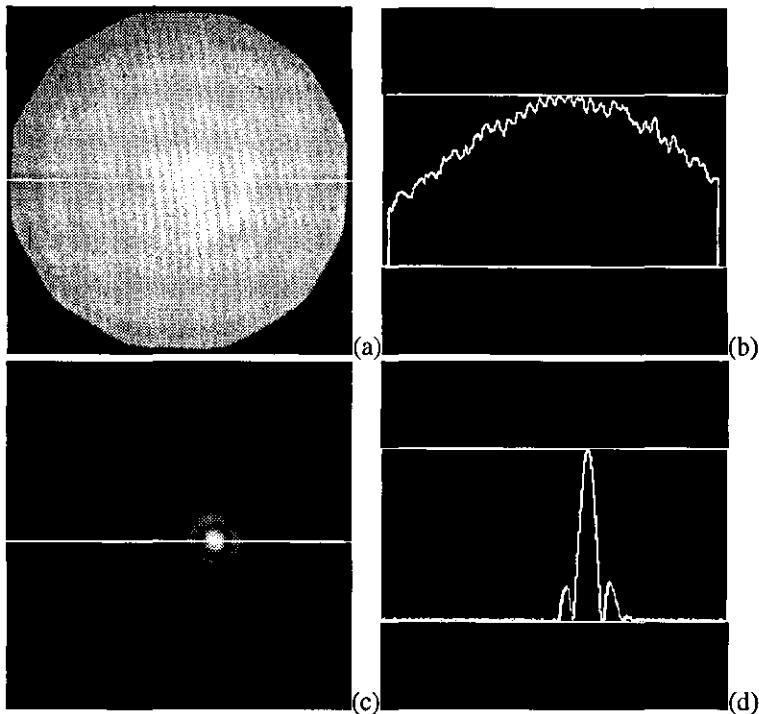


Fig. 5.28 - Image and diffraction functions for the glass slide. (a)- image magnitude; (b)- profile marked in (a); (c)- diffraction quasi magnitude; (d)- profile marked in (c).

All the results obtained in this set were quite consistent. Figure 5.29(a) shows one of these recovered wrapped phase distributions when the normalized root mean squared error was 2.73×10^{-3} after 200 iterations.

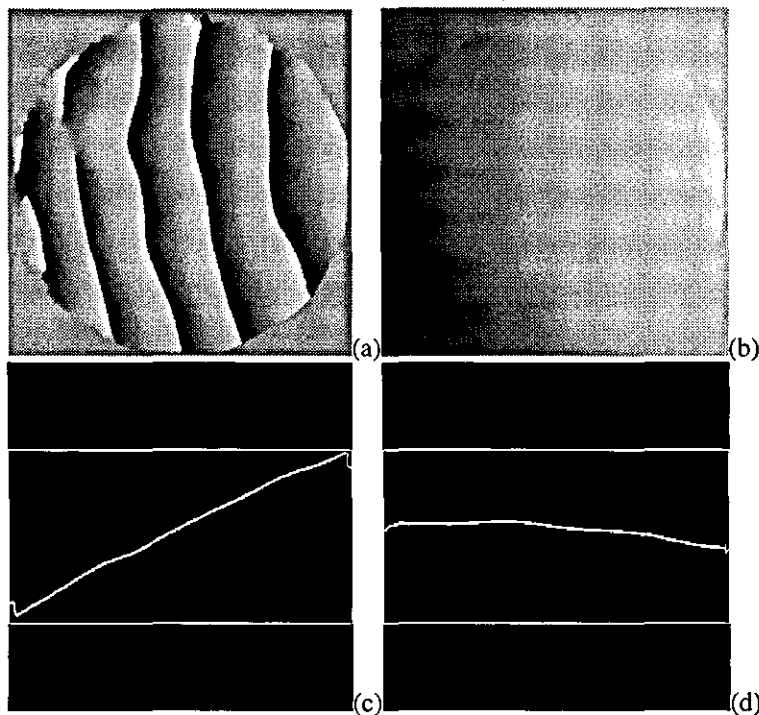


Fig. 5.29 - Recovered phase function for the glass slide. (a)- relative distribution; (b)- absolute distribution; (c)- middle horizontal profile of (b); (d)- middle vertical profile of (b).

After 200 more iterations the error was reduced to 9.18×10^{-4} . The corresponding unwrapped phase distribution is shown in Fig. 5.29(b). Also, in Fig. 5.29(c) and (d) are shown respectively the middle horizontal and vertical profiles of the unwrapped phase distribution. The absolute phase changes measured along these directions correspond to approximately 10π rad (5λ) and 2π rad (λ) respectively.

To give a better idea about the absolute phase distribution and consequently about the shape of the glass slide surface a pseudo three-dimensional representation of the unwrapped phase is shown in Fig. 5.30. This unwrapped phase representation corresponds to the left-hand side view of Fig. 5.29(b) and consequently the bottom of the image corresponds to the left-hand side of the glass slide.

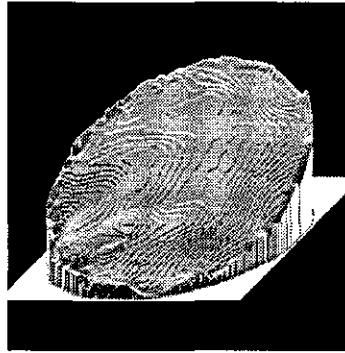


Fig. 5.30 - Pseudo 3D representation of the unwrapped phase distribution for the glass slide.

The glass slide used in this experiment was a microscope glass slide approximately 1 mm thick. Its thickness was experimentally measured with a micrometer in four different points in the considered circular area of the slide. The thickness measurements made near the top and bottom of the circular area of the slide gave respectively, $1,045 \mu\text{m}$ and $1,043 \mu\text{m}$. The difference in thickness from top to bottom is about $2 \mu\text{m}$. The same measurements have been made on the left and right sides and the results were $1,042 \mu\text{m}$ and $1,049 \mu\text{m}$ for the thickness at the left and right sides respectively. The difference in thickness for these measurements is then about $7 \mu\text{m}$.

The objective now is, to show and prove that the results obtained from the retrieved phase distribution can be used to calculate the difference in thickness over the glass slide as well as the angle of the deviated light in a particular direction.

Since the glass slide can be seen to act as a prism in the field due to the approximately linear phase changes, one can also calculate from the data the angle of this equivalent prism. Consider the schematic diagram of a prism which is illuminated by a plane parallel beam of monochromatic light as shown in Fig. 5.31.

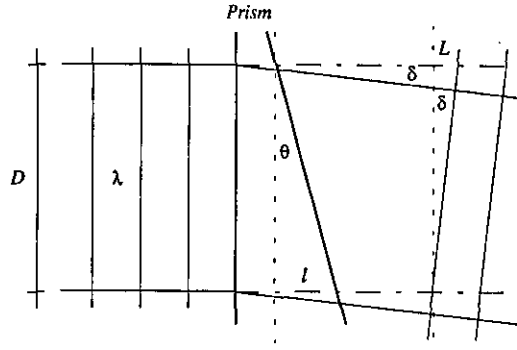


Fig. 5.31 - Deviation of light by a prism.

Assuming that the thickness of the glass surface changes linearly as in Fig. 5.31, one can relate the angle θ of the prism with the angle δ of the deviated light by the following equation [4]

$$\delta = (n - n_o)\theta, \quad (5.4-9)$$

where $n = 1.5$ is the refractive index of the glass and $n_o = 1$ is the refractive index of the air. From Fig. 5.31 one can express for a small angle δ that

$$\tan \delta \approx \delta = L/D, \quad (5.4-10)$$

where $D = 16$ mm is the diameter of the considered aperture, and $L = N_p \lambda$ represents the distance equivalent to the number of phase changes N_p with wavelength $\lambda = 632.8$ nm. Recalling the phase change results previously obtained, one has for the horizontal direction a value $L_h \approx 5\lambda = 3.164 \mu\text{m}$ and for the vertical direction the value $L_v \approx \lambda = 0.6328 \mu\text{m}$, which gives using Eq. (5.4-10), the deviation angle of $\delta_h = 197.75 \mu\text{rad}$ in the horizontal direction and the deviation angle of $\delta_v = 39.55 \mu\text{rad}$. Now using the Eq. (5.4-9), one can obtain the angle of the equivalent prism which has in the horizontal direction an angle $\theta_h = 2\delta_h = 395.5 \mu\text{rad}$ and for the vertical direction an angle $\theta_v = 2\delta_v = 79.1 \mu\text{rad}$. To confirm these results, one can use in this specific case and by coincidence an interferometric approach. A careful observation of the object image (see Fig. 5.28(a)) reveals some interference fringes over the aperture. This assumes that the fringes are due to interference between the transmitted and internally reflected beam in the glass slide. The spacing between the fringes is not constant, but if it is assumed to be constant, one can calculate the change in thickness of the glass Δl associated with each fringe using the well known interferometric relation [6]

$$\Delta l = \frac{\lambda/2}{n}, \quad (5.4-11)$$

which gives for the values of $\lambda = 632.8$ nm and $n = 1.5$, $\Delta l = 0.21$ μm . In this way and knowing the number of fringes in both the horizontal and vertical directions it will be possible to establish the corresponding difference in thickness for the two directions. The number of fringes in the horizontal direction is approximately $N_h = 30$ and in the vertical direction $N_v = 6$. This gives the difference in thickness equal to $l_h = \Delta l N_h = 6.3$ μm for the horizontal direction and the value $l_v = \Delta l N_v = 1.26$ μm for the vertical direction. Comparing these values with the values measured with the micrometer, which were 7 and 2 μm respectively, one can consider the agreement to be quite good. To complete the calculations the values l_h and l_v can be used to evaluate the prism angle θ and the deviation angle δ along the corresponding directions. Referring to Fig. 5.31, since

$$\tan \theta \approx \theta = \frac{l}{D} \quad (5.4-12)$$

for small angles, one obtains the values $\theta_h = l_h/D = 393.75$ μrad and $\theta_v = l_v/D = 78.75$ μrad for the horizontal and vertical directions respectively. Again, a comparison between these values and the values previously calculated, which were $\theta_h = 395.5$ μrad and $\theta_v = 79.1$ μrad , shows good agreement. For the above reasons and due to the consistency and reproducibility of the retrieved phase distributions in all the iterative processes one can consider this cycle of phase retrieval a success. Successful results with similar conclusions have also been obtained with a 3 mm thick perspex slide.

Soap Film

In this example a thin water/soap film was made by means of a 20 mm diameter wire ring immersed and then withdrawn from a soap solution. Held vertically in the object plane the film drained and the different images were recorded simultaneously before the film ruptured. Figure 5.32 depicts the recorded data associated with the distribution functions taken from the image and Fourier domains on the experimental arrangement when the thin soap film was used as an object disturbance. In particular, the object magnitude function obtained from CCD_1 is shown in Fig. 5.32(a) and the corresponding profile in Fig. 5.32(b). The quasi magnitude function of the diffraction pattern taken from CCD_2 is shown in Fig. 5.32(c) and in (d) its horizontal profile.

A set of procedures to retrieve the phase distribution in the image domain, for the vertical soap film, has been executed. This set was composed by repeating a 200 iteration process several times feeding the algorithm with the magnitude functions previously calculated using the steps described in the beginning of this section.

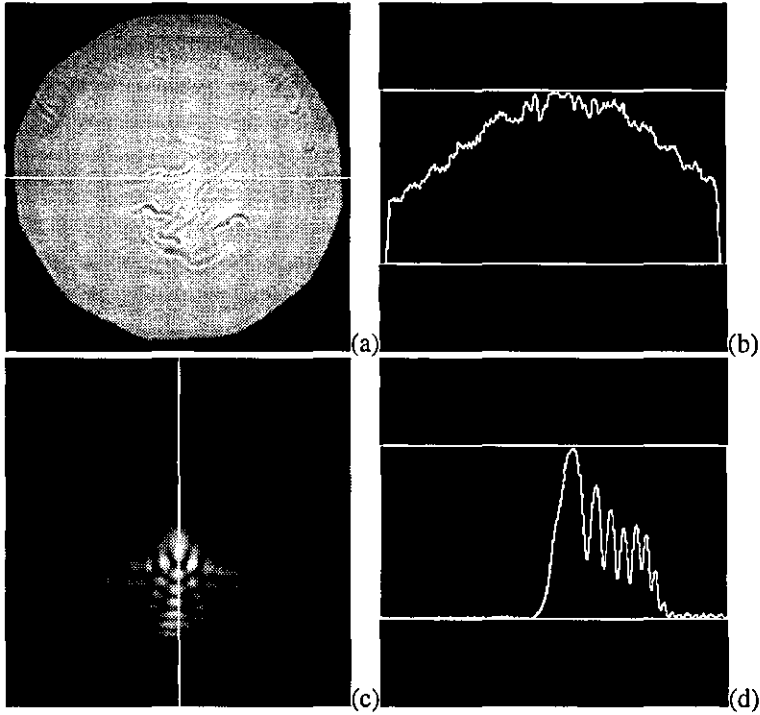


Fig. 5.32 - Image and diffraction functions for the soap film. (a)- image magnitude; (b)- profile marked in (a); (c)- diffraction quasi magnitude; (d)- profile marked in (c).

All the results obtained in this set were quite consistent. Figure 5.33(a) shows one of these recovered wrapped phase distributions when the normalized root mean squared error was 3.85×10^{-3} after 200 iterations.

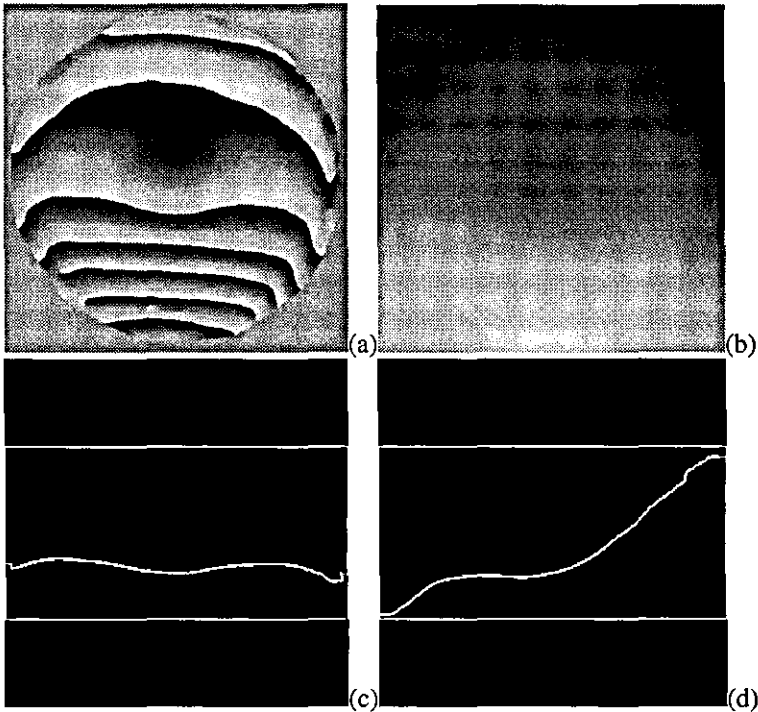


Fig. 5.33 - Recovered phase function for the soap film. (a)- relative distribution; (b)- absolute distribution; (c)- middle horizontal profile of (b); (d)- middle vertical profile of (b).

After 200 more iterations the error was reduced to 1.52×10^{-3} . The corresponding unwrapped phase distribution is shown in Fig. 5.33(b). Also, in Fig. 5.33(c) and (d) are shown respectively the middle horizontal and vertical profiles of the unwrapped phase distribution. The absolute phase changes measured along these directions correspond to approximately 16π rad (8λ) and π rad ($\lambda/2$) respectively. As one can observe from the above pictures and profiles, the phase changes follow a clear non-linear distribution along the horizontal and vertical directions.

To give a better idea about the shape of the surface of the soap film the absolute phase distribution can be used to generate a pseudo three-dimensional representation as it is shown in Fig. 5.34. This unwrapped phase representation corresponds to the top view of Fig. 5.33(b) and consequently the bottom of the image corresponds to the top of the vertical soap film.

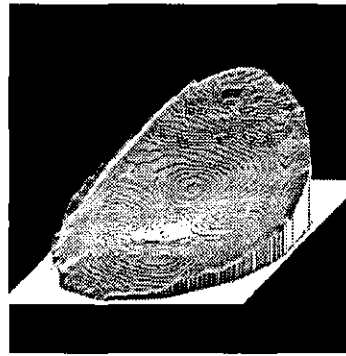


Fig. 5.34 - Pseudo 3D representation of the unwrapped phase distribution for the soap film.

One can consider this cycle of phase retrieval a success due to the consistency and reproducibility of the retrieved phase distributions in all the iterative processes.

Heat Flow

Convective heat flow above a hot body represents a more realistic situation often viewed by schlieren photography. Figure 5.35 depicts the recorded data associated with the distribution functions taken from the image and Fourier domains on the experimental arrangement when a hot soldering iron tip was used to produce a heat flow disturbance in the object domain OD . The iron tip was partially located in the aperture field. In particular, the object magnitude function obtained from CCD_1 is shown in Fig. 5.35(a) and the corresponding profile in Fig. 5.35(b). The quasi magnitude function of the diffraction pattern taken from CCD_2 is shown in Fig. 5.35(c) and in (d) its horizontal profile. It is interesting to note that no disturbance is visible in the image magnitude (Fig. 5.35(a)) as may be expected for large heat flows above a candle flame for example.

A set of procedures to retrieve the phase distribution in the image domain, for this real object, has been executed. This set was composed by repeating a 200 iteration process several times feeding the algorithm with the magnitude functions previously calculated using the steps described in the beginning of this section.

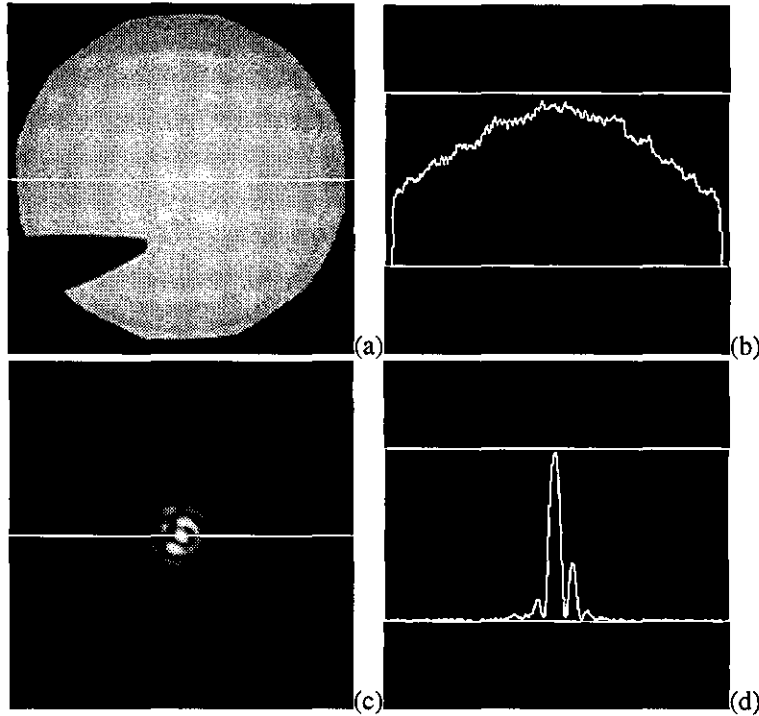


Fig. 5.35 - Image and diffraction functions for the heat flow. (a)- image magnitude; (b)- profile marked in (a); (c)- diffraction quasi magnitude; (d)- profile marked in (c).

The results obtained in this set were quite interesting because of a randomly apparent difference between them which suggests an inconsistency in the recovered phase distribution. Fig. 5.36(a) shows one of these recovered wrapped phase distributions when the normalized root mean squared error was 5.73×10^{-3} after 200 iterations. Also, in Fig. 5.36(b) is shown another recovered wrapped phase distribution in which the normalized root mean squared error was 4.36×10^{-3} for the same number of iterations. After 200 more iterations in each of the referred retrieval processes, the error was reduced respectively to 3.29×10^{-3} and 3.47×10^{-3} .

In fact these phase distributions seem to be different, but since they are relative phase distributions when the absolute phase distribution is calculated, one can observe that they represent exactly the same function, as shown by the unwrapped phase distribution in Fig. 5.36(c). For that reason, the results obtained from the phase retrieval process should always be judged after the unwrapping phase operation to avoid wrong conclusions. In mathematical and computational terms both the distributions, the wrapped and the unwrapped, represent the same and unique function.

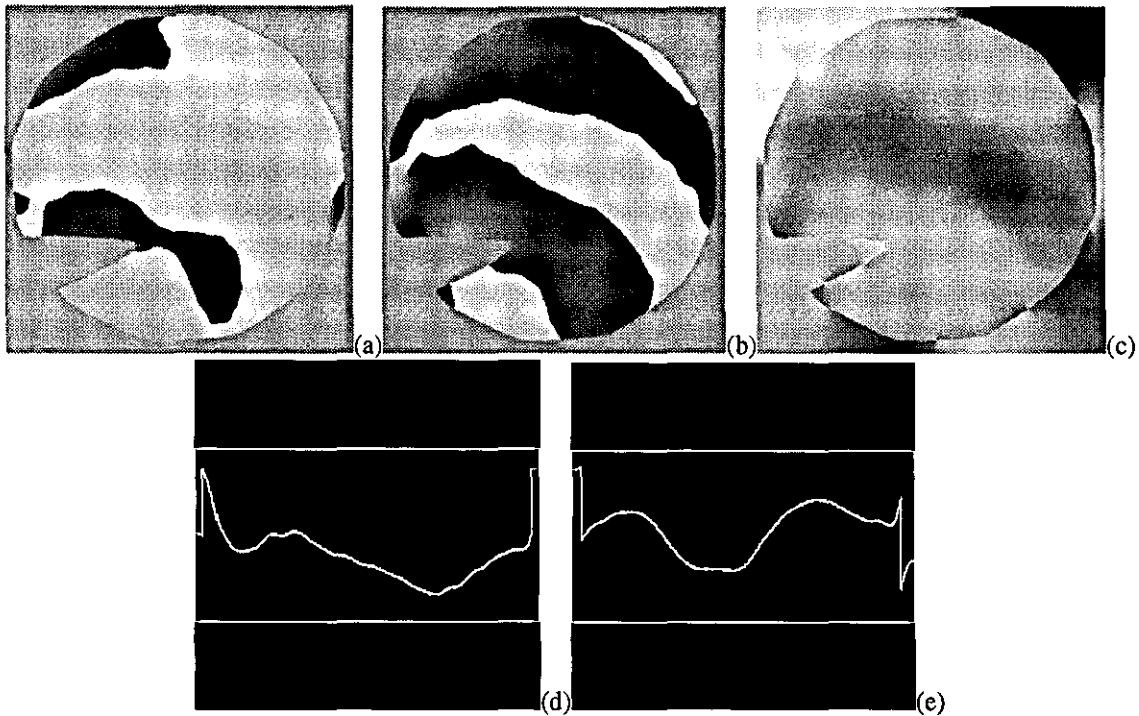


Fig. 5.36 - Recovered phase functions for the heat flow. (a)- relative distribution *A*; (b)- relative distribution *B*; (c)- absolute distribution *A* and *B*; (d)- middle horizontal profile of (c); (e)- middle vertical profile of (c).

Also, in Fig. 5.36(d) and (e) are shown respectively the middle horizontal and vertical profiles of the unwrapped phase distribution. The absolute phase changes measured along these directions correspond to approximately 2π rad (λ) and π rad ($\lambda/2$) respectively, and the maximum phase change being approximately 4π rad (2λ).

To give a better idea about the absolute phase distribution a pseudo three-dimensional representation of the unwrapped phase is shown in Fig. 5.37. This unwrapped phase representation corresponds to the left-hand side view of Fig. 5.36(c).

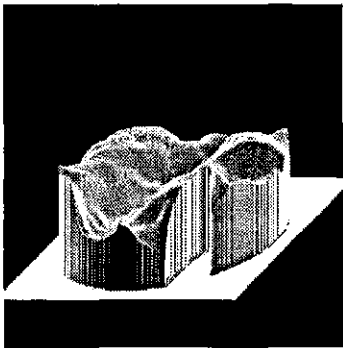


Fig. 5.37 - Pseudo 3D representation of the unwrapped phase distribution for the heat flow.

One can consider this cycle of the phase retrieval a success due to the consistency and reproducibility of the retrieved phase distributions in all the iterative processes.

No Object Disturbance

Figure 5.38 depicts the recorded data associated with the distribution functions taken from the image and Fourier domains on the experimental arrangement when no object disturbance exists in the object domain OD . In particular, the magnitude function taken from CCD_1 is shown in Fig. 5.38(a) and the corresponding profile in Fig. 5.38(b). The quasi magnitude function of the diffraction pattern of the single aperture taken from CCD_2 is shown in Fig. 5.38(c) and in (d) its horizontal profile.

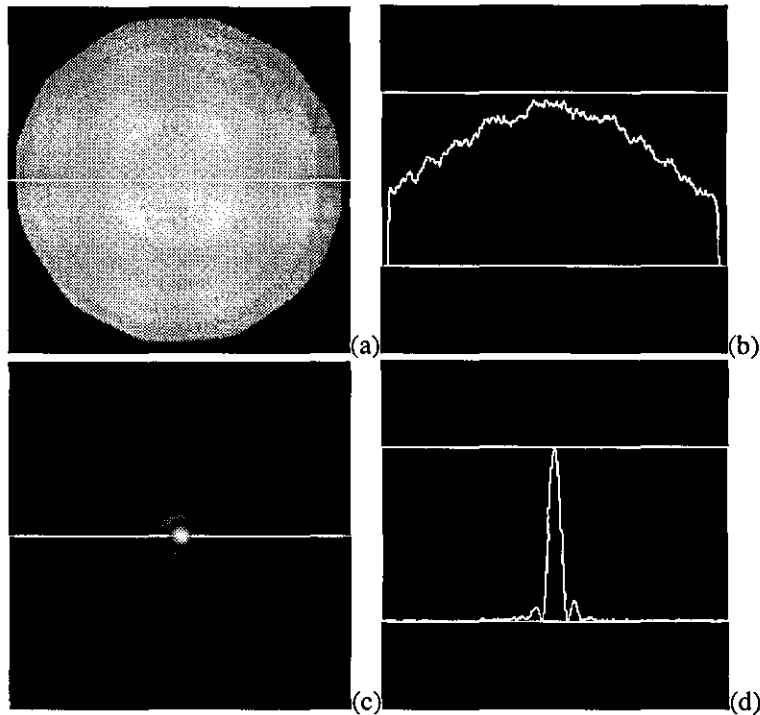


Fig. 5.38 - Image and diffraction functions for the single aperture (No Object). (a) image magnitude; (b)- profile marked in (a); (c)- diffraction quasi magnitude; (d)- profile marked in (c).

A set of procedures to retrieve the phase distribution in the image domain have been executed. This set was composed by repeating a 200 iteration process several times feeding the algorithm with the magnitude functions previously calculated as before.

All the results obtained in this set were quite consistent. Figure 5.39(a) shows one of these recovered wrapped phase distributions when the normalized root mean squared error was 4.21×10^{-3} after 200 iterations. After 200 more iterations the error was reduced to 1.73×10^{-3} . Due to the very small phase changes ($< 2\pi$ rad) the corresponding unwrapped phase distribution coincides with the relative phase distribution shown in Fig. 5.39(a). Also, in Fig. 5.39(b) and (c) are shown respectively the middle horizontal and vertical profiles of the unwrapped phase distribution.

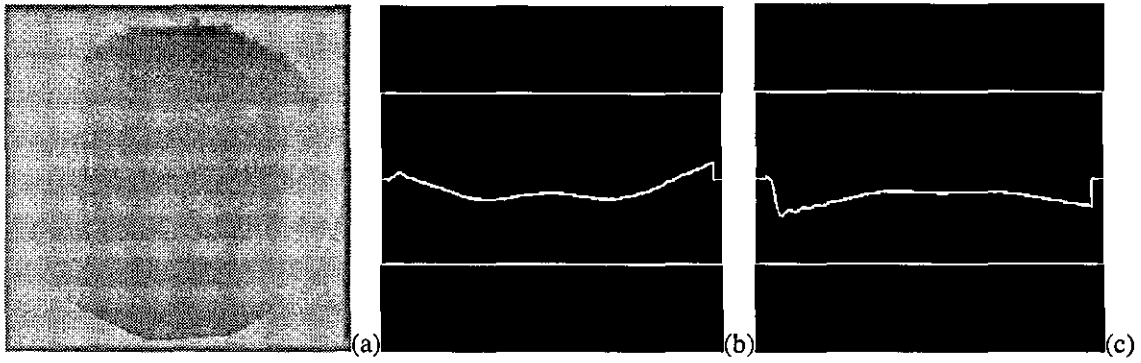


Fig. 5.39 - Recovered phase function for the single aperture (No Object). (a)- absolute distribution; (b)- middle horizontal profile of (a); (c)- middle vertical profile of (a).

As one can observe from the above picture and profiles, for this case the phase change measured for both the horizontal and vertical directions is approximately $\pi/2$ rad ($\lambda/4$) which is indeed a very small phase change. This information can be used to study the wave-front in terms of deformations which have been generated by imperfections in the lens system. In this particular case, the Fourier lens used on the experimental apparatus is a high quality lens. Therefore, one can expect and interpret the following results, as a wave-front distortion due to imperfections in the collimating lens. Such imperfections result in aberration effects which distort the desired plane wave-front.

This information can also be used to verify the quality of the laser beam on the experimental apparatus where a plane parallel beam of monochromatic light is required. From the above result one can observe that the wave-front has an almost constant phase distribution over the circular aperture and consequently it can be considered a plane wave-front.

The shape of the wave-front can be represented in a pseudo three-dimensional display using the recovered phase distribution, as shown in Fig. 5.40. The phase representation corresponds to the left-hand side view of Fig. 5.39(a) and consequently the bottom of the image corresponds to the left-hand side of the aperture.

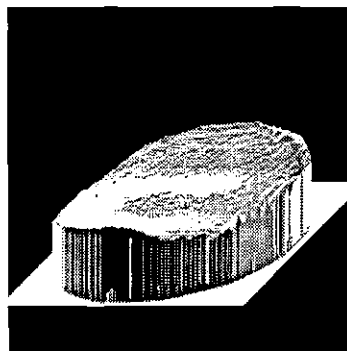


Fig. 5.40 - Pseudo 3D representation of the phase distribution for the wave-front.

In order to verify the sensitivity of this process, the collimating lens L_1 on the experimental apparatus, shown in Fig. 5.17, has been slightly displaced from its actual location to introduce some distortion on the laser beam wave-front. In this way, the beam can not now be considered a plane parallel beam of light and one wishes to recover the phase distribution associated with this new wave-front.

Figure 5.41 depicts the recorded data associated with the distribution function in the Fourier domain in the experimental arrangement after the collimating lens had been displaced slightly and no object disturbance is placed in the object domain OD . In particular, the magnitude function taken from CCD_1 is the same shown in Fig. 5.38(a) and the corresponding profile in Fig. 5.38(b). The quasi magnitude function of the diffraction pattern of the single aperture taken from CCD_2 is shown in Fig. 5.41(a) and in (b) its horizontal profile.

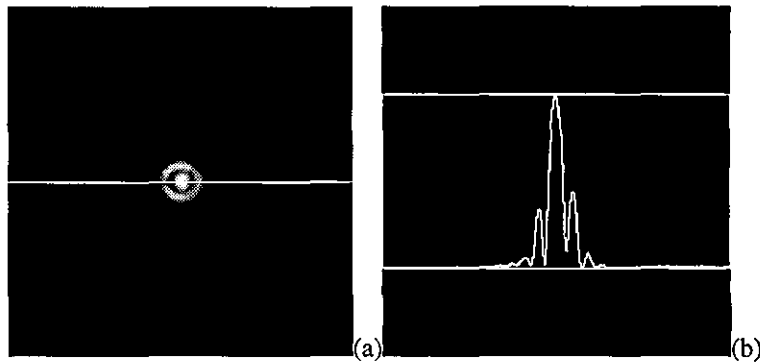


Fig. 5.41 - Diffraction functions for the single aperture (Lens Displaced and No Object).

(a)- diffraction quasi magnitude; (b)- profile marked in (a).

Another set of procedures to retrieve the phase distribution in the image domain, for this new wave-front, was executed. This set was composed by repeating a 200 iteration process several times feeding the algorithm with the magnitude functions previously calculated in the same way as before.

All the results obtained in this new set were quite consistent. Figure 5.42(a) shows one of these recovered wrapped phase distributions when the normalized root mean squared error was 6.58×10^{-3} after 200 iterations. After 200 more iterations the error was reduced to 1.19×10^{-3} . Once again, due to the very small phase changes ($< 2\pi$ rad) the corresponding unwrapped phase distribution coincides with the relative phase distribution shown in Fig. 5.42(a). Also, in Fig. 5.42(b) and (c) are shown respectively the middle horizontal and vertical profiles of the absolute or unwrapped phase distribution.

For this particular case, the phase changes were still very small but the phase distribution is not constant as in the previous example. The result reflects the effect of the lens displacement and consequently, one can evaluate qualitative and quantitatively the wave-front of concern. This makes this procedure an important intermediate step in the experimental set up to obtain a plane parallel beam of monochromatic light. For this reason, one can consider this procedure as a calibration process to be performed on the experimental arrangement before carrying out any object experiment.

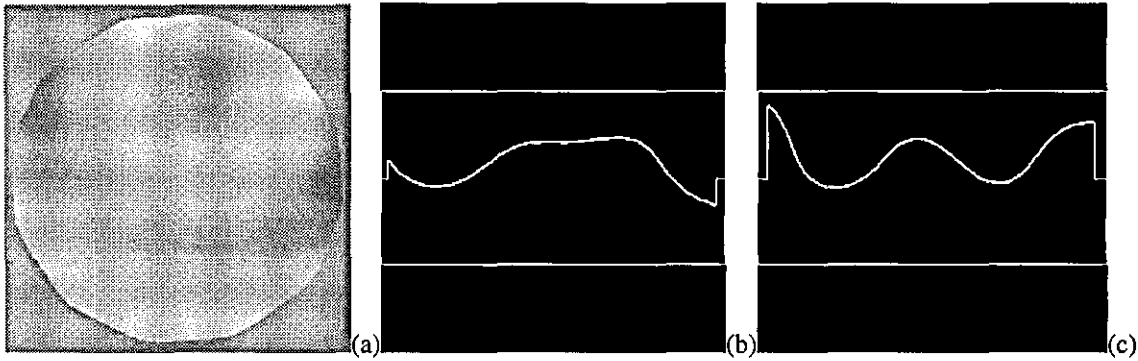


Fig. 5.42 - Recovered phase function for the single aperture (Lens Displaced/No Object).

(a)- absolute distribution; (b)- middle horizontal profile of (a); (c)- middle vertical profile of (a).

Figure 5.42(a) shows exactly what would be expected if the collimating lens is not precisely at the focal distance from the spatial filter pinhole. Inspection of this image also shows it to be the same as an interferogram where an off axis spherical wave-front interferes with a plane reference wave. The shape of the wave-front can be represented in a pseudo three-dimensional display using the recovered phase distribution, as it is shown in Fig. 5.43. This phase representation corresponds to the left-hand side view of Fig. 5.42(a) and consequently the bottom of the image corresponds to the left-hand side of the aperture.

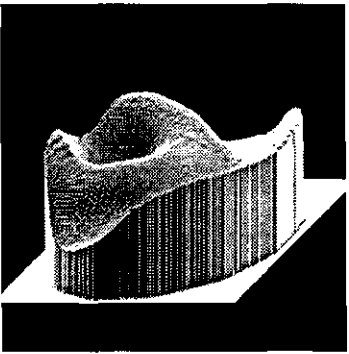


Fig. 5.43 - Pseudo 3D representation of the phase distribution for the new wave-front.

One can consider this cycle of phase retrieval a success due to the consistency and reproducibility of the retrieved phase distributions in all the iterative processes.

5.4.5 Conclusions

The present real experiments proved that the iterative algorithm for phase retrieval can be used successfully to recover the phase distributions associated with an optical object disturbance from accurate measurements of the intensity and the quasi magnitude information of its image and diffraction pattern taken from an experimental apparatus. The experimental arrangement used in the experiments carried out has been fully described and it enables the measurement of the intensity distribution in the image domain as well as the quasi magnitude distribution in the Fourier domain of an optical wave-front disturbance.

The scale relation established between the image and Fourier domains has been carefully analysed in order to define the limits of the system based on the practical limitations of the apparatus and also in results that have been obtained in the previous simulations. It has been found in the simulations that a phase change limit of approximately $12 \times 2\pi$ rad must be imposed. Since the wave length λ of the laser radiation used as a source of monochromatic light in the apparatus can be expressed in terms of 2π phase steps, the phase change limit for this particular experimental arrangement is then twelve wave lengths (12λ). In fact this is a rather narrow band of phase changes for different sorts of object disturbances. For this reason, it is only possible to recover phase distributions of optical object disturbances within this range, but in principle the system with the appropriate modifications should be able to recover any phase change distribution. The modifications concern simply the size of the sampled data arrays, bearing in mind that large array sizes imply a considerably increase in the processing time. For a 1024×1024 pixel array the processing time increases by a factor of 5. Consequently 200 iterations of the algorithm would take 2.5 hours instead of 30 minutes which is the time required for a 512×512 pixel array. In some cases time is not a problem, the problem is then related with the configuration capabilities of the computer in order to enable the manipulation of very large arrays of information, which is in general an impediment.

Accuracy of measurements, in particular in the Fourier domain, is the key to the success of the phase retrieval process. This system uses image sensors that have the capability to perform electrically an operation called the gamma correction. This feature has been investigated and it has been found possible to measure instead of the usual intensity information the quasi magnitude information when the image detectors are set to perform a gamma correction of 0.45. In this way, measurements of low light intensities can be performed and more accurate measurements of the corresponding magnitude can

be achieved. The success in all the retrieval phase distributions that have been carried out for different real object disturbances is due to the accuracy in the measured information obtained from such image sensors. The experimental results obtained were consistent for both the linear and non-linear phase changes considered, and their reproducibility makes this method of phase retrieval an important tool for optical wave-front reconstruction of object disturbances.

Another important achievement is that this method can be used as a calibration procedure for optical beams when no object is placed in the object domain. This calibration procedure concerns the correct positioning of the optical components, in particular the collimating and the Fourier lens in the experimental arrangement, in order to obtain as a source of illumination a plane parallel beam of monochromatic light. As a consequence of the limited phase change that can be detected with the system the sensitivity of this method is reasonably high, being possible to detect changes in the phase in the order of $2\pi/20$ rad, or $\lambda/20$.

5.5 Summary

In this chapter the iterative algorithm for phase retrieval has been used to recover the phase distributions of object disturbances. Firstly, a set of computer simulation have been carried out to analyse the performance and the success of the phase retrieval process and a set of conditions for the correct convergence to the right solution have also been established. Secondly, the simulated results have been extended to experimental results.

The consistency and reproducibility in both the simulated and real results makes this method for the phase retrieval problem an important tool to be used in research of optical wave-fronts of object disturbances and in its reconstruction. The recovery of the phase distribution gives to the researcher all the necessary information about the object disturbance and mathematical models of complex functions can be recreated for the wave-front of concern in both the image and Fourier domains of an optical system. The method has its limitations in practice, but in principle it can be applied successfully to any kind of object disturbance.

Another way to look at the system is to consider its potential as a single beam interferometer. The experiments without any object disturbance highlights this possibility. The final result in all the experiments is a phase distribution map which has

been obtained using a single path through the disturbance (or indeed no disturbance) together with the application of the phase retrieval algorithm. All that is requested is a stored reference beam and its diffraction pattern together with the Fourier diffraction image and the corresponding disturbed real image. The limitation which has been identified is concerned with the maximum phase change or rather the maximum gradient of the phase change which can be accommodated by the imaging sensors.

References

1. R. W. Gerchberg and W. O. Saxton, "A Practical Algorithm for the Determination of Phase from Image and Diffraction Plane Pictures", *Optik* **35**, 237-246 (1972).
2. D. C. Ghiglia and L. A. Romero, "Robust Two-dimensional Weighted and Unweighted Phase Unwrapping that uses Fast Transforms and Iterative Methods", *J. Opt. Soc. Am. A* **11**, 107-117 (1994).
3. R. N. Bracewell, *The Fourier Transform and its Applications*, 2nd Edition, McGraw-Hill Ed., New York (1986).
4. F. A. Jenkins and H. E. White, *Fundamentals of Optics*, 4th Edition, McGraw-Hill Ed., Singapore (1981).
5. Hecht, *Optics*, 2nd Edition, Addison-Wesley Publishing Company, Inc., (1987).
6. R. D. Guenther, *Modern Optics*, John Wiley & Sons, Inc., New York (1990).
7. J. W. Goodman, *Introduction to Fourier Optics*, Reissued Edition, McGraw-Hill, Inc., New York (1988).
8. H. C. Andrews and B. R. Hunt, *Digital Image Restoration*, Prentice-Hall, Inc. Englewood Cliffs, New Jersey (1977).

CHAPTER 6

Computer Generation of Schlieren Images

6.1 Introduction

Many problems of science and engineering involve substances that are transparent and non luminous, so that their observation by direct visual or photographic methods is difficult. The phenomena that are of interest frequently involve changes of the refractive index across the field to be investigated, which can then be visualized by using optical techniques that depend on the effects of the refractive index change on the transmission of the light as described in chapter one of this thesis. Several techniques are available but this chapter is specially concerned with the schlieren technique.

The schlieren technique depends on the deflection of a ray of light from its undisturbed path when it passes through a medium in which there is a component of the gradient of refractive index normal to the ray. It may be shown [1,2] that the deviation angle of the ray is proportional to the refractive index gradient in the direction normal to the ray.

In Fig. 6.1 is basically sketched a typical schlieren apparatus. Normally a knife edge K can be manipulated in the focal plane of the lens L_2 , usually referred to as the Fourier domain FD , to modify the field of view in the image formation process. In this way, a set of schlieren images can be obtained on the screen, usually referred to as the image domain ID , in which the gradual intensity change of the light corresponds to changes in the angle of the deviated light by an object disturbance Q located in the object domain OD .

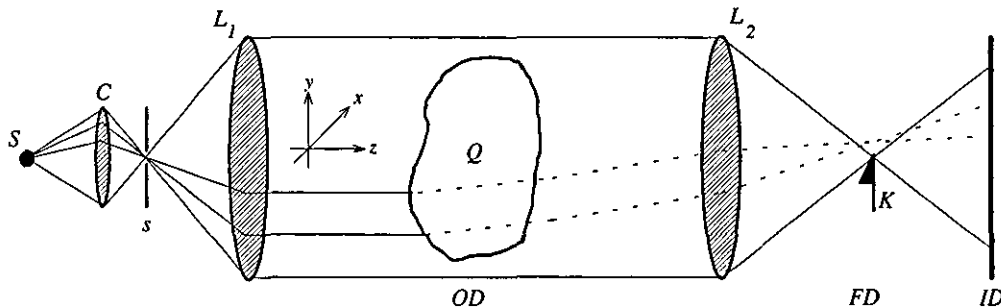


Fig. 6.1 - The arrangement of a typical schlieren apparatus.

In order to quantitatively characterize the object disturbance the position of the knife edge is extremely important. Since the light can be deviated in any direction the evaluation of the deviation angle is not always easy and a large set of schlieren images is

needed. An alternative to the knife edge is the use of a circular stop where the deviated light in all directions has a contribution to the schlieren image.

The problem of concern is related to the position of the knife edge or to the size of the circular stop which are in practice quite critical. If for example, the refractive index gradient is small the deviation or disturbance of the diffraction pattern in the Fourier domain is also very small and positional accuracy of the order of the Airy disc is necessary. When the object disturbance is stable in time and space there is no problem to generate such schlieren images with the corresponding practical limitations. But, if the object disturbance is not stable, the reproducibility of such an event will be impossible and there is no correlation between two or more of such schlieren images.

Knowing both the magnitude and phase information of a complex diffraction pattern function in the Fourier domain associated with an optical wave-front produced by an object disturbance it will be possible to recreate on the computer schlieren images for different knife-edge positions or for different sizes of the circular stop. This new technique for computer generation of schlieren images involves:

- the recreation of a discrete complex wave-front function associated with an object diffraction pattern, as described in chapter five;
- the simple manipulation of a *computer knife edge or a circular stop* to remove, that is, set to zero, specific components (magnitude and phase) in the recreated discrete function;
- and finally, the use of this modified function as input to the discrete inverse Fourier transform algorithm, which has been described in chapter three, to generate the so called schlieren image.

In the next sections several results obtained with this new technique for both simulated and real experiments will be presented and discussed as well as the advantages and disadvantages of such a technique over the conventional schlieren technique.

6.2 Wave-Front Reconstruction and Schlieren Images

As shown in chapter five, the Gerchberg and Saxton iterative algorithm can successfully be used to recover phase information from records of intensity and quasi-magnitude taken from the image and the Fourier domains of an optical system. For this case in particular, one is concerned with the retrieved phase in the Fourier domain in order to

recreate on the computer the discrete complex valued function that represents the complex diffraction pattern of an object disturbance. Later this discrete function will be modified on the computer as if it had been modified by a real knife-edge or circular stop in an optical system in order to generate on the computer screen a schlieren image by inverting the Fourier transformation of the resulting discrete diffraction pattern function.

6.2.1 Simulated Example

This section is devoted to a presentation and discussion of a simulated example of an optical wave-front reconstruction associated with an acoustic transient (shock wave) in water. Later, using the reconstructed wave-front function of the event a set of computer schlieren images will be generated on the computer screen. The simulated results for different knife-edge positions and for a circular stop will be interpreted in terms of practical realization.

Acoustic Transients (Shock Waves) in Water

The experimental arrangement normally used to obtain quantitative information about an acoustic field produced by an electromagnetic acoustic transducer [3] is shown in Fig. 6.2. The resulting pressure variations produced in water, which will be simulated here, can be imaged in the image domain and quantitatively analyzed by manipulation of knife-edges in the Fourier domain.

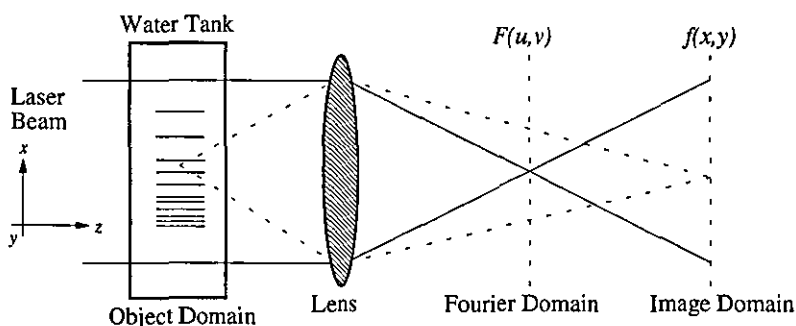


Fig. 6.2 - Schematic simulation/experimental arrangement.

A change of pressure in a fluid leads to a change in density and a corresponding change in refractive index. When a plane parallel beam of monochromatic light is used to illuminate such an object disturbance, one can expect changes in the phase of the wave-front that arrives from the object disturbance due to the refractive index changes. It was deduced and shown in [3] the existence of an expression that relates the changes in

refractive index with the changes in pressure produced by a planar shock wave in water. The relation can be expressed by

$$\frac{\partial n}{\partial P} = \frac{n_o - 1}{v^2 \rho_o}, \quad (6.2-1)$$

where ∂n represents the change in refractive index, ∂P the change in pressure, n_o and ρ_o the refractive index and density of the medium at a certain standard temperature and pressure, respectively, and v the speed of the sound in the medium. Evaluating Eq. (6.2-1) for $n_o = 1.33$, $\lambda = 694\text{nm}$, $v = 1,490\text{m/s}$, and $\rho_o = 998\text{m}^3\text{Kg}$ gives $\partial n/\partial P = 1.5 \times 10^{-4} (\text{MPa})^{-1}$. Recent studies [4,5] have shown that the variation of a shock wave pressure $P(t)$ with respect to time, can be modelled by the following equation

$$P(t) = 2P_m \exp[-\alpha t] \cos\left(\omega t + \frac{\pi}{3}\right) K(1 - \exp[-\beta t]), \quad (6.2-2)$$

where P_m is the peak pressure, α is the decay constant, $\omega = 2\pi f$ is the radial frequency, and the factor $K(1 - \exp[-\beta t])$ allows the selection of non-zero rise times. For example, a 100 ns rise time can be selected by choosing $\beta = 4.6 \times 10^7 \text{ s}^{-1}$, and $K = 1.03$. Substituting these values and those for a shock wave with a peak pressure of 0.5 MPa, a period of 6 μs ($\omega = 1.0467 \times 10^6 \text{ rads/s}$) and a decay constant of (for a high damped response) $\alpha = 9.1 \times 10^5 \text{ s}^{-1}$, into Eq. (6.2-2) gives the pressure wave-form shown in Fig. 6.3. A similar model can be obtained with respect to space, $P(x)$, simply making $x = vt$, which gives a spatial extent of shock, for the period $\Delta t = 6 \mu\text{s}$, equal to $\Delta x = 8.94 \text{ mm}$.

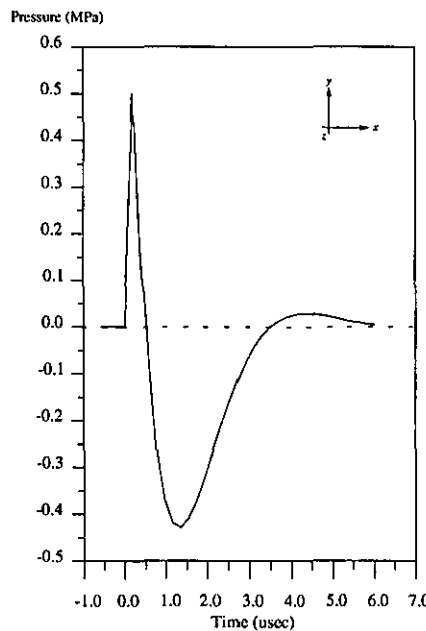


Fig. 6.3 - The modelled shock wave pressure.

Assuming that the shock wave travels in the negative x direction (see Fig. 6.2), then the pressure distribution in any plane parallel to the yz plane is constant and follows the distribution as indicated above in Fig. 6.3. It should be noted that the practical finite extent in both the y and z directions has been fixed in this simulation to $\Delta y = \Delta z = 10\text{mm}$.

Using the pressure distribution modelled by Eq. (6.2-2) and recalling Eq. (6.2-1) one can model the refractive index distribution n_i in the medium as

$$n_i = (n_o + \partial n) = (1.33 + 1.5 \times 10^{-4} \partial P), \quad (6.2-3)$$

which is drawn in Fig. 6.4(a). For the first peak pressure $\partial P = +0.5$ MPa and the corresponding refractive index will be $n_i = 1.330075$, similarly for the second peak pressure $\partial P = -0.429$ MPa which corresponds to a refractive index $n_i = 1.329936$.

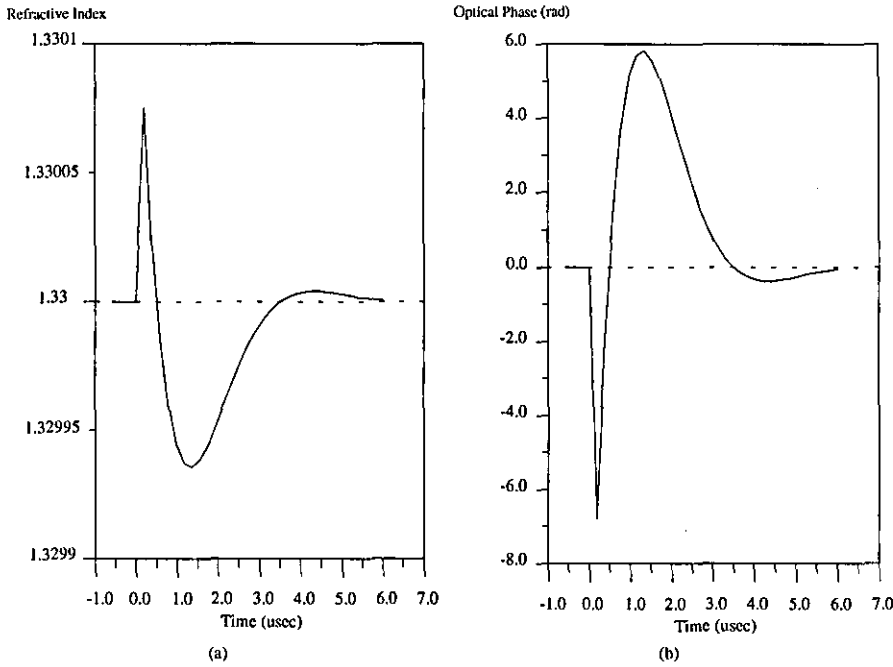


Fig. 6.4 - The modelled refractive index change (a), and (b), the optical phase change.

Since the shock wave disturbance is illuminated perpendicularly to the xy plane, as shown in Fig. 6.2, one can now calculate the optical path changes in the direction of the light propagation (z direction) according to the following equation

$$\Delta z_i = n_i \Delta z, \quad (6.2-4)$$

where Δz_i is the optical path length associated with the spatial length Δz with refractive index n_i . If the optical path length changes then the phase of the plane wave-front that

travels through the disturbance will change accordingly. The phase change distribution can be modelled using the equation

$$\Delta\theta_i = (\Delta z_{ref} - \Delta z_i) \frac{2\pi}{\lambda}, \quad (6.2-5)$$

where $\Delta\theta_i$ is the phase change associated with the optical path Δz_i with respect to the optical reference path Δz_{ref} , which is shown in Fig. 6.4(b). This phase distribution will be used in the following simulation where the main purpose is to show that it will be possible to recover the simulated phase distribution, using the algorithm described in chapter four, from magnitude measurements in the image and Fourier domains of the simulated optical system shown in Fig. 6.2. Then it will be possible to generate on the computer screen a set of schlieren images using the reconstructed discrete complex valued diffraction pattern function.

Assuming that the object disturbance is coherently illuminated through a circular aperture of 16 mm in diameter and the light intensity distribution on the aperture has a Gaussian profile, one can calculate its magnitude distribution and then combine it with the modelled phase to create on the computer the complex valued wave-front function arriving from the object as a 512×512 discrete array. These dimensions are chosen here because they match the practical values which have been used in real experiments.

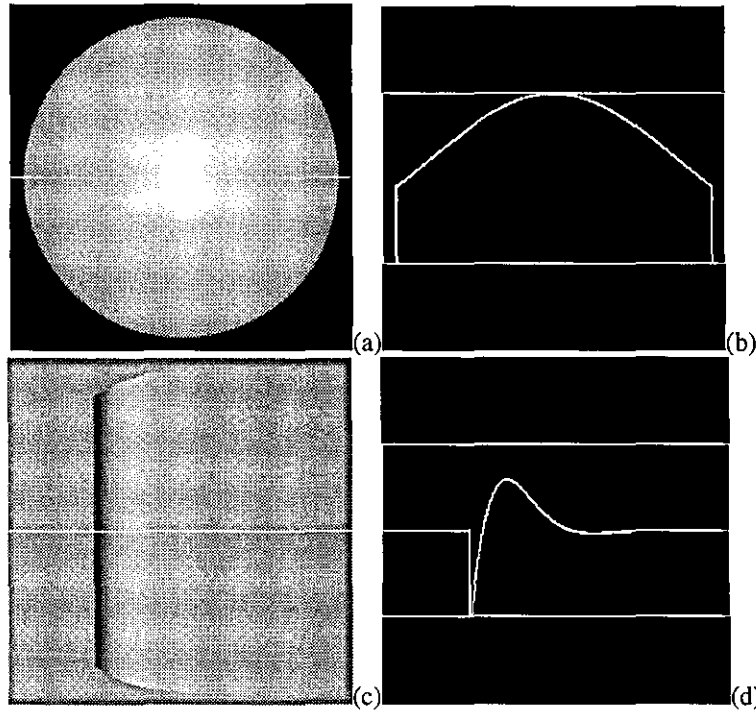


Fig. 6.5 - Simulated object wave-front function. (a)- magnitude distribution; (b)- profile marked in (a); (c)- phase distribution; (d)- profile marked in (c).

In particular, Fig. 6.5(a) shows the two-dimensional magnitude distribution of the illuminating light in the object or image domain and the corresponding horizontal profile in Fig. 6.5(b). Also, in Fig. 6.5(c) and (d) are shown the phase distribution associated with the shock wave in the object domain and its horizontal profile, respectively. The phase scale is confined to the interval $[-2\pi, +2\pi]$, as reported in Fig. 6.4(b), and the object disturbance will be considered as a pure phase-object because only the optical phase is affected.

Now, using the direct Fourier transformation algorithm this simulated function can be transformed to obtain the corresponding complex diffraction pattern function of the shock wave object. The result is shown in Fig. 6.6, where the magnitude of the diffraction pattern 20x magnified is shown in (a) and the marked profile in (b).

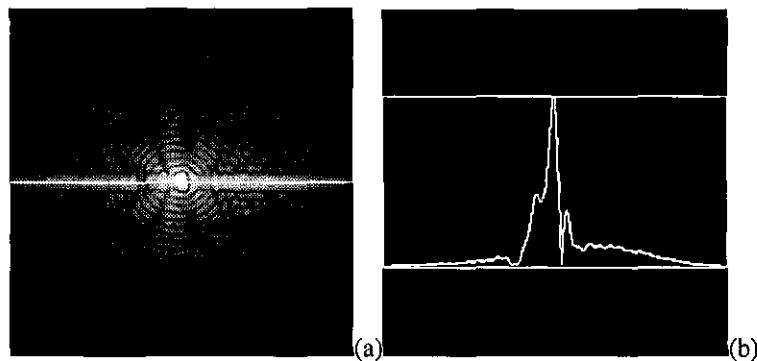


Fig. 6.6 - Simulated Fourier diffraction wave-front function. (a)- magnitude distribution;
(b)- profile marked in (a).

According to the procedures described in chapter five, a set of iterations has been executed to retrieve the phase information in the object or image domain using the Gerchberg and Saxton algorithm. The input data to the algorithm were the discrete magnitude distribution functions for both the image and Fourier domains as they are shown in Figs. 6.5(a) and 6.6(a).

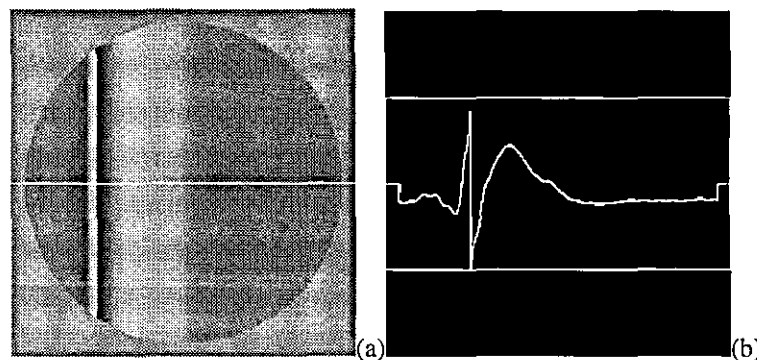


Fig. 6.7 - Recovered simulated object phase function. (a)- phase distribution;
(b)- profile marked in (a).

After 200 iterations the normalized root mean square error, as it has been defined in chapter four Section 4.4.2, was reduced to 2.54×10^{-3} and the resulting recovered phase distribution for the image domain is shown in Fig. 6.7. Figure 6.7(a) depicts the recovered phase distribution and Fig. 6.7(b) the horizontal profile marked in (a). The results obtained in other iterative processes were all consistent with the original phase function which is a guarantee of the convergence of the algorithm to the correct solution.

One can now use this recovered object phase function combined with the corresponding object magnitude function to calculate the discrete complex diffraction pattern function by direct Fourier transformation. The generation of schlieren images on the computer screen, as it was mentioned before, concerns the modification of this discrete complex diffraction pattern function by means of a computer knife-edge or circular stop, before the inverse Fourier transformation to obtain the resulting image of the object disturbance. To make clear these ideas, Fig. 6.8 shows the resulting magnitude distributions in the Fourier domain when a knife-edge is applied to remove (set to zero) the left-half and the right-half of the Fourier field, Figs. 6.8(a) and (b) respectively, and a circular stop is applied to the centre region of the Fourier field, Fig. 6.8(c).

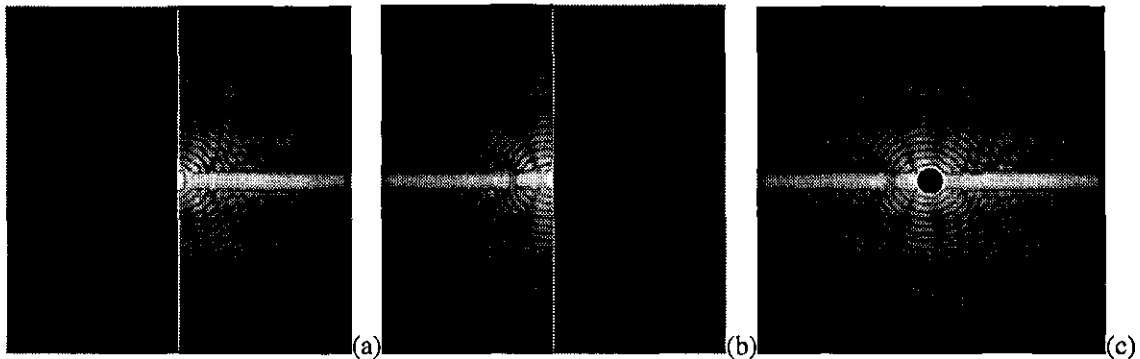


Fig. 6.8 - The effect of a computer knife-edge and a circular stop on the diffraction pattern function.

(a)- left-half; (b)- right-half; (c)- central region.

From those pictures, one can observe that some magnitude and phase components were removed from the diffraction pattern function and therefore they do not contribute to the image formation process when the inverse Fourier transformation is performed. In other words, with the knife-edge located on the left-hand side, only the light deflected to the right by the object disturbance, which corresponds to positive gradients of the refractive index, has a contribution for the resulting schlieren image, as shown in Fig. 6.9(a). Similarly, with the knife-edge located on the right-hand side, only the light deflected to the left by the object disturbance, which corresponds to negative gradients of the refractive index, has a contribution for the resulting schlieren image, as shown in Fig.

6.9(b). Finally, with the circular stop located on the centre, only the low order deflection of the light is completely removed and the deflected light in all the directions has a contribution to the resulting schlieren image, as shown in Fig. 6.9(c). Since in this particular case, the big deflections occur mainly to the left-hand side and to the right-hand side due to the refractive index distribution, the result obtained with the circular stop could be seen as the superposition of the results obtained with the left-half and right-half knife-edges.

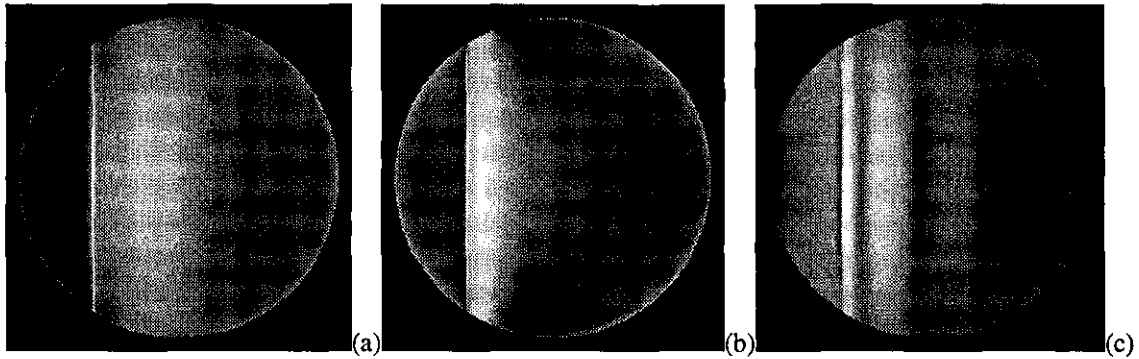


Fig. 6.9 - The computer generated schlieren images of the simulated shock wave. Knife-edge position: (a)- left-half; (b)- right-half. Circular stop position: (c)- central region.

As mentioned before, the diffraction pattern function in the Fourier domain is magnified 20x. This magnification factor appears, as discussed in chapter five, to preserve the scale relation between the Fourier and image domains. As a result of this magnification, one can get on the computer a 20x more accurate position of the knife-edge or sizing of the circular stop in the diffraction pattern. For example, the knife-edge can be positioned with an accuracy of 1 pixel which corresponds to $8.5 \mu\text{m}$ on the scale of the pictures shown in Fig. 6.8. To obtain in practice the equivalent positioning, an accuracy of $\approx 425 \text{ nm}$ would need to be achieved. For the circular stop shown in Fig. 6.8(c), which is a circle 40 pixels in diameter, it corresponds to a circle, $340 \mu\text{m}$ in diameter. To obtain in practice the equivalent schlieren picture under the same conditions of this simulation an opaque circular stop $\approx 17 \mu\text{m}$ in diameter would need to be placed in the Fourier domain of the optical system.

In Fig. 6.10 is shown, as a comparison, a schlieren image of a real planar shock wave travelling in water. This image has been taken from a schlieren experiment, where the shock wave was produced by an electromagnetic acoustic transducer [3,6], and a circular stop 3 mm in diameter was used to remove the low order frequency components in the Fourier domain. The first and second peak pressures correspond to approximately $+5.0 \text{ MPa}$ and -3.0 MPa respectively, and the physical dimensions of the disturbance were $\Delta y = \Delta z = 60 \text{ mm}$ and $\Delta x = 8.94 \text{ mm}$. This means that the optical phase changes

in the wave-front are confined to the interval $[-130\pi, +110\pi]$ which is approximately 60x bigger than the one considered in the simulation. Due to the increased pressure, the refractive index changes are larger and the angle of deviated light is also large. A 14 mm displacement of the diffraction pattern have been measured which corresponds to a deviation angle of ≈ 2.75 mrad or $\approx 1.57^\circ$. For this reason the system sensitivity and the sizing of the circular stop (3 mm) are not so critical, as described in the simulation (17 μ m), to obtain a good visualization of the shock wave, where the similarities with the simulated schlieren image are clear.



Fig. 6.10 - The experimental schlieren image of a real shock wave using a circular stop.

As mentioned in chapter five, the algorithm and the developed system has a limitation on the maximum phase change that can be detected. Consequently it will be extremely difficult with the present system to recover on the computer the phase distribution for the experimental shock wave due to the phase change involved. Since the simulation was a 60x de-magnification of the real experiment it can be used as a means to predict experimental results. On the other hand, the practical limitations on the accurate positioning of a knife-edge and on the sizing of a circular stop make this technique of computer generation of schlieren images very attractive for the study and visualization of very weak optical disturbances.

6.2.2 Real Examples

This section is now devoted to present and discuss real examples of an optical wave-front reconstruction associated with a thin soap film and with a weak heat flow produced by a soldering iron. The wave-front reconstruction process recalls some of the results obtained and discussed in the previous chapter. Using these reconstructed wave-front functions a set of computer schlieren images will be generated for different knife-edge positions and for different circular stop sizes. The results will be interpreted in terms of their practical realization.

Thin Soap Film

Figure 6.11 shows the magnitude functions calculated from the recorded intensities taken from the image and Fourier domains on the experimental arrangement, which has been described in detail in chapter five and schematically shown in Fig. 6.2, when a thin soap film was used as an object disturbance. The soap film was coherently illuminated by a plane parallel beam of light through a circular aperture of 16 mm in diameter and these magnitude functions have been stored in two 512×512 discrete arrays.

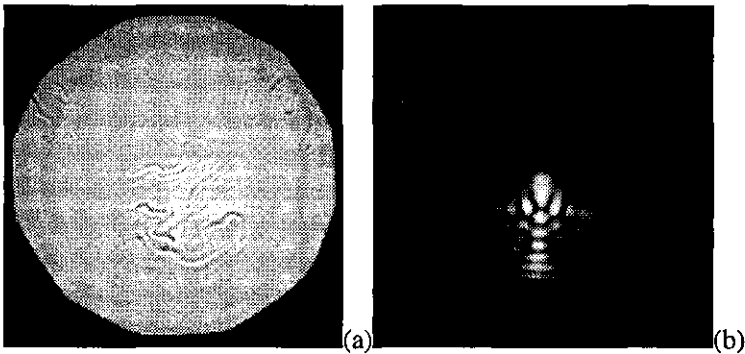


Fig. 6.11 - Real object magnitude functions. (a)- object magnitude;
(b)- Fourier diffraction magnitude.

Using the result obtained in chapter five, where these two magnitude functions have been used to recover the phase in the image domain, it is possible to recreate on the computer the corresponding discrete complex diffraction pattern function.

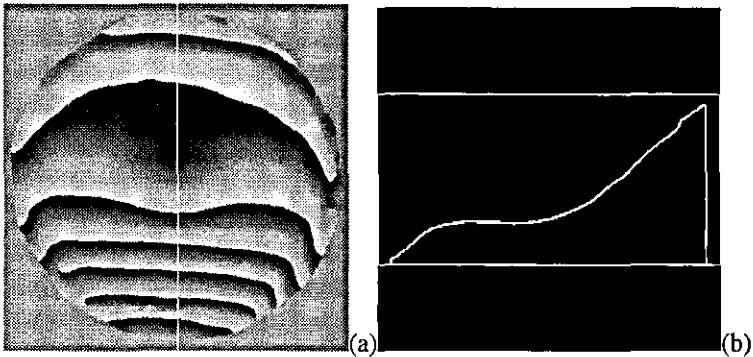


Fig. 6.12 - Recovered real object phase function. (a)- relative phase distribution;
(b)- unwrapped profile marked in (a).

To obtain that function, the object magnitude distribution, shown in Fig. 6.11(a), should be combined with the recovered object phase distribution, shown in Fig. 6.12(a), to form a discrete complex image function which is then transformed using the direct Fourier transform algorithm into a discrete diffraction pattern function. Now, this recreated diffraction pattern function can be modified any time and a set of computer schlieren

images can be generated. As an example, Fig. 6.13(a) shows the intensity distribution of the computer generated schlieren image when a circular stop was used to remove the low order frequency components in the complex diffraction pattern function of this real object before the image formation process. Figure 6.13(b) shows one vertical profile of the intensity distribution, where the deviation of the light due to the refractive index gradient associated with the recovered phase change, see Fig. 6.12(b), is clearly visible.

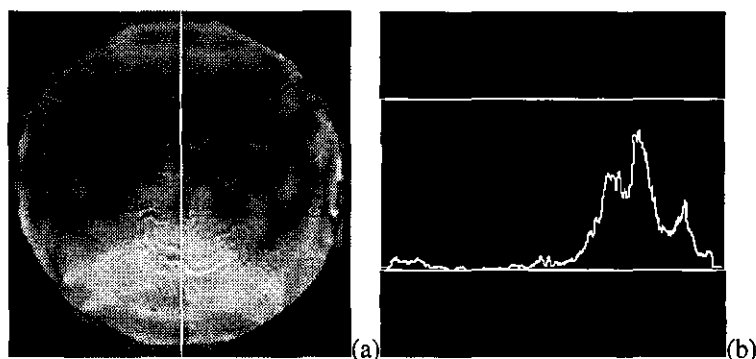


Fig. 6.13 - The computer generated schlieren image of the thin soap film. (a)- image intensity distribution; (b)- profile marked in (a).

The circular stop used was 80 pixels in diameter which corresponds to a circle with 680 μm in diameter for the actual scale in the Fourier domain. Experimentally it was found impossible to produce the equivalent schlieren image by manipulation of a real circular stop. According to the 20x scale factor involved in the experiment the diameter of an opaque circular stop would need to be $\approx 34 \mu\text{m}$ in this particular application.

Heat Flow from a Hot Iron

Figure 6.14 shows the magnitude functions calculated from the recorded intensities also taken from the image and Fourier domains on the same experimental arrangement, described in detail in chapter five and schematically shown in Fig. 6.2.

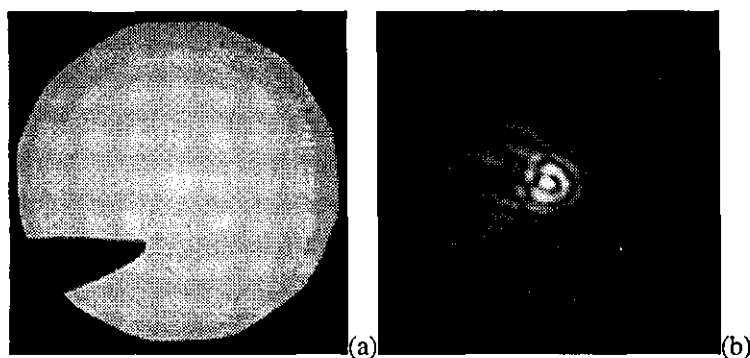


Fig. 6.14 - Real object magnitude functions. (a)- object magnitude; (b)- Fourier diffraction magnitude.

Here a hot piece of iron was used to produce a field disturbance in the object domain. The soldering iron tip was partially placed in the aperture field which was coherently illuminated by a plane parallel beam of light through the same circular aperture, 16 mm in diameter.

These magnitude functions have been stored in two 512×512 discrete arrays in order to be used in the object phase retrieval process. Using the two results obtained in the phase recovery process, presented and discussed in the previous chapter, it is possible to recreate on the computer the two corresponding discrete complex diffraction pattern functions.

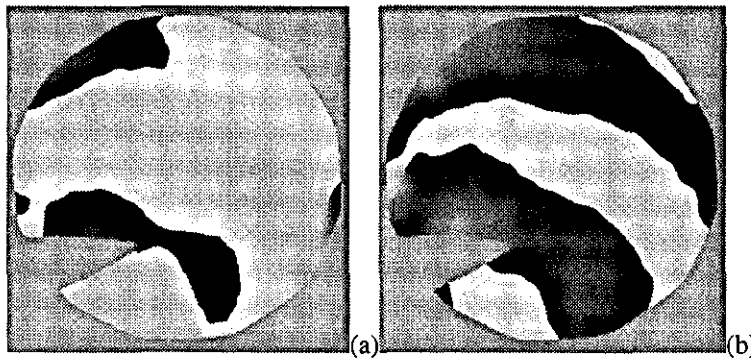


Fig. 6.15 - Recovered real object phase functions. (a)- relative phase distribution A;
(b)- relative phase distribution B.

To obtain these functions, the object magnitude distribution, shown in Fig. 6.14(a), should be combined with the recovered object phase distribution A, shown in Fig. 6.15(a), to form a discrete complex image function which is then transformed using the direct Fourier transform algorithm into a discrete diffraction pattern function. The object magnitude distribution, shown in Fig. 6.14(a), should be combined with the other recovered object phase distribution B, shown in Fig. 6.15(b), to form a discrete complex image function which is then transformed using the direct Fourier transform algorithm into another discrete diffraction pattern function. As mentioned in chapter four, the phase recovery process recovers the relative phase distribution and not the absolute phase distribution. The relative phase is the absolute phase wrapped into the range $[-\pi, +\pi]$ and 2π jumps occur for phase variations of more than 2π . For this reason the relative phase distributions shown in Fig. 6.15, which have been obtained in two different retrieval processes, seen to be different but in reality they represent exactly the same absolute phase distribution if an algorithm for phase unwrapping [7,8] is used to obtain the corresponding absolute phases, as shown in the previous chapter. The following results show, using a different approach, that the recovered phase distributions are the same.

First, and using the recreated diffraction pattern function obtained from the relative phase distribution A , a set of computer schlieren images has been generated to show the effect of different knife-edge positions on the visualization of the heating flow in the field. Figure 6.16 shows four schlieren images obtained on the computer for left-half, right-half, top-half, and bottom-half knife-edge positions on the Fourier domain. From these pictures one can gain information about the refractive index gradients, produced by the hot piece of iron, respectively to the right, left, bottom, and top directions because these pictures are the result of the deflected light in these directions.

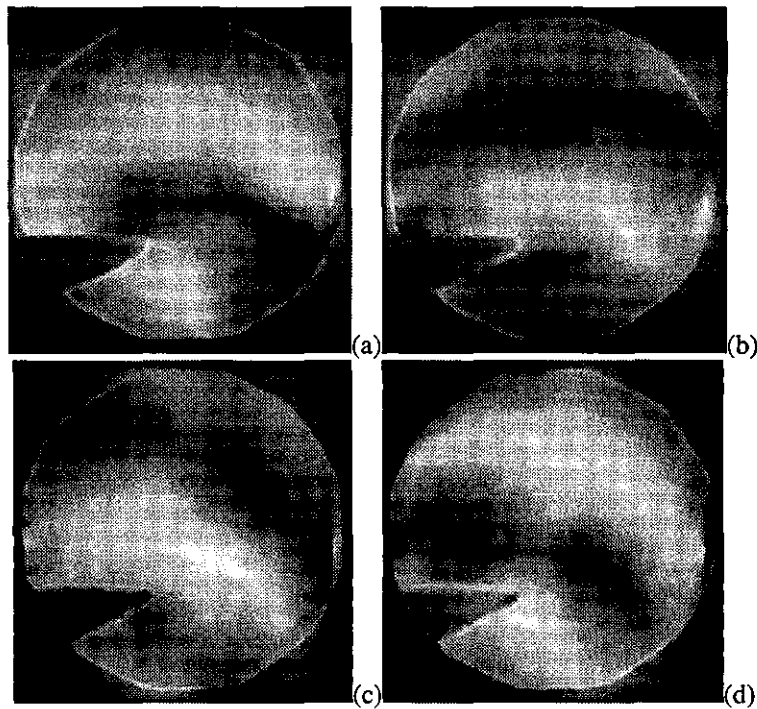


Fig. 6.16 - The computer generated schlieren images of the heat flow. Knife-edge position: (a)- left-half; (b)- right-half; (c)- top-half; (d)- bottom-half.

Figure 6.17 shows the result of the computer schlieren image when a circular stop with 30 pixels in diameter ($\approx 255 \mu\text{m}$) was placed in the central region of the Fourier domain.

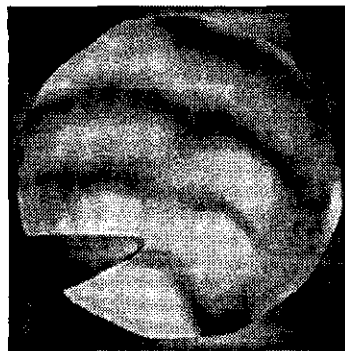


Fig. 6.17 - The computer generated schlieren image of the heat flow. Circular stop position at the central region.

This result can be seen as a superposition of the four different results shown in Fig. 6.16. Using the recreated diffraction pattern function obtained from the relative phase distribution B , another set of computer schlieren images has been generated to show the effect of different circular stop sizes on the visualization of the heating flow in the field. Figure 6.18 shows three schlieren images obtained on the computer for a circular stop with 10 pixels ($\approx 85 \mu\text{m}$), 20 pixels ($\approx 170 \mu\text{m}$), and 30 pixels ($\approx 255 \mu\text{m}$) in diameter, placed on the central region of the Fourier domain.

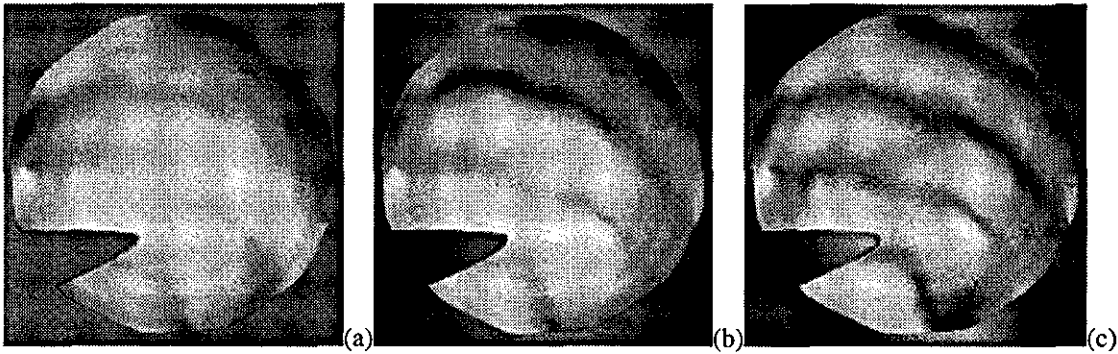


Fig. 6.18 - The computer generated schlieren image of the heat flow. Circular stop diameter: (a)- 10 pixels; (b)- 20 pixels; (c)- 30 pixels.

From those pictures one can observe that if the diameter of the circular stop is too small the resulting schlieren image does not differ much from the original image because only a few low order frequency components were removed from the image formation process and consequently the visualization of the heating flow is really poor. To obtain a good visualization the diameter of the circular stop should be increased to remove more low order frequency components in order to visualize only the deflected light in all the directions but making a specific angle with the optical axis. Figure 6.18(b) and (c) shows the results for a 10 pixel increment in the diameter of the stop, which can be compared with the result shown in (a). Good visualization does not mean bigger diameters of the circular stop. If the diameter is increased too much and the optical disturbance does not deflect the light to bigger angles, which is equivalent to say small phase changes, the result is a completely dark image because most of the important frequency components in the Fourier domain were removed and only the very high frequency components, with very little information, make a contribution to the schlieren image.

The computer schlieren images generated using the two recreated discrete diffraction pattern functions with a circular stop 30 pixels ($\approx 255 \mu\text{m}$) in diameter are shown in Figs. 6.17 and 6.18(c). The similarities between these two schlieren images are obvious which proves that the two recovered relative phase distributions correspond to the same

absolute phase distribution and also that the phase unwrapping procedure is not absolutely necessary in this process of generation of schlieren images but it will be necessary if the objective is the display of a phase distribution map.

Experimentally it was found impossible to produce the equivalent schlieren images by manipulation of a real circular stop in the experimental arrangement. The diameter of an opaque circular stop would need to be $\approx 12 \mu\text{m}$ in this particular case, due to the 20x scale factor involved in the experiment, to make possible the practical reproduction of such schlieren images. Also mentioned in the previous chapter, when the optical phase changes are very small, approximately 2λ for this particular experiment, the changes in the diffraction pattern distribution are confined to a very small area located in the focal plane of the imaging lens (Fourier domain). Consequently due to the small optical phase changes the light suffers small deflections and an accurate sizing and positioning of the circular stop $< 20 \mu\text{m}$ would be required to obtain adequate sensitivity in the selection of the frequency components to go through the optical system to produce the equivalent schlieren images.

It should be noted that due to the accurate sizing and positioning of the circular stop, the schlieren images generated can reveal an interesting effect which is related to the high sensitivity of this technique. The effect is usually called the diffraction or interference due to the edges of the opaque object used as the circular stop or knife edge [2]. In practice the visualization of diffraction bands in schlieren images, such as the ones shown in Figs. 6.17 and 6.18(c), is not easy unless a high sensitivity schlieren instrument is used. In this way, this technique also makes possible the study of such an effect, in particular in situations where the amount of optical phase change is small, avoiding the use of complex and expensive schlieren instruments.

Recalling the result obtained for the schlieren image of the soap film, shown in Fig. 6.12(a), one can observe that no diffraction bands are visible in the schlieren intensity distribution. For this case the optical phase change was approximately 8λ (see chapter five in Section 5.4.4), which is bigger than the phase change in the current example (2λ). The difference is that the size of the circular stop used was quite big making the diffraction effect very weak and consequently not visible on the schlieren image.

To make this idea clear, in Fig. 6.19 is shown the result of a computer generated schlieren image of the heating flow, when the diameter of the circular stop was set to 50 pixels ($\approx 425 \mu\text{m}$) which corresponds in practice, to a circular stop of $\approx 21.25 \mu\text{m}$.

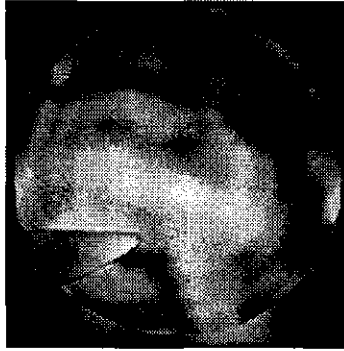


Fig. 6.19 - The computer generated schlieren image of the heat flow.

The resulting schlieren image does not show the diffraction bands as before and also, the information concerning the heat flow is now closer to the information that can be obtained from the schlieren images shown in Fig. 6.16. This technique of computer generation of schlieren images seems to complement the conventional schlieren on the visualization of very weak effects and currently some research to study the diffraction effects on the generated schlieren images is under way.

6.3 Conclusions

A new technique for computer generation of schlieren images has been proposed and described. The technique involves the recreation of a discrete complex wave-front function, associated with an object diffraction pattern, using the retrieved phase distribution. The modification of this function by removing specific components, using a computer knife-edge or a circular stop, enables the generation of schlieren images on the computer screen. The schlieren images are simply the results of the inverse Fourier transformation of the modified diffraction pattern functions and remarkably high sensitivities can be obtained.

Quantitative schlieren measurements are very difficult to perform and the simplicity of the described technique makes possible the generation of schlieren images on the computer screen as many times as required. This means that the actual experiment is performed only once to acquire and record on the computer the data associated with an event. After that, everything will be performed on the computer. The technique has a limitation on the maximum phase change across the field that can be recovered, which is approximately $12 \times 2\pi \equiv 12\lambda$, due to the scale factor between the image and Fourier domains and the limiting size of the arrays used to record the information, but in principle the technique will work for any phase change when the necessary detector and computer requirements are satisfied. Also, the high sensitivity of the technique to study

weak effects associated with small phase changes makes it a useful complement to the conventional schlieren and shadow techniques which are normally used to study quantitatively strong disturbance effects which involve large phase changes or large refractive index gradients.

The simulation of an optical wave-front reconstruction, associated with an acoustic transient (shock wave) in water, has shown the capabilities of the developed digital image processing system to estimate the results before the practical realization of the experiment if a proper model is used. The reconstructed wave-front has been used to generate a set of schlieren images which show the agreement of the results obtained with the expected results from the theory. The simulated results also show how accurate the positioning of a knife-edge or the sizing of a circular stop would need to be to generate experimentally the equivalent schlieren images. The practical limitations on the accurate positioning or sizing make this technique very attractive for result prediction and estimation.

Simulation is not all that can be done with this system. It has also shown the application of the technique to real experimental situations. Results for a thin soap film and for a heating flow produced by a soldering iron have been presented and discussed in terms of their practical realization. Due to the small optical phase changes involved in the considered experiments and due to the corresponding practical limitations, the generation of schlieren images for these object disturbances was only possible on the computer screen. The application of this technique to more complex object disturbances may be more difficult but it will certainly complement the conventional schlieren and shadow methods.

References

1. D. W. Holder and R. J. North, *Schlieren Methods*, Her Majesty's Stationery Office, London (1963).
2. L. A. Vasil'ev, *Schlieren Methods*, Keter Inc., New York (1971).
3. M. T. Carnell and D. C. Emmony, "Quantitative Schlieren Measurements of a High Energy Electromagnetic Transducer Acoustic Shock Field", *Optics & Laser Technology* **26**, 385-391 (1994).
4. C. C. Church, "A Theoretical Study of Cavitation Generated by an Extracorporeal Shock Wave Lithotripter", *J. Acoust. Soc. Am.* **86**, 215-227 (1989).
5. A. Philipp, M. Delius, *et. al.*, "Interaction of Lithotripter-Generated Shock Waves with Air Bubbles", *J. Acoust. Soc. Am.* **93**, 2496-2509 (1993).
6. M. T. Carnell, R. D. Alcock, and D. C. Emmony, "Optical Imaging of Shock Waves Produced by a High Energy Electromagnetic Transducer", *Physics Medicine and Biology* **38**, 1575-1588 (1993).
7. D. C. Ghiglia and L. A. Romero, "Robust Two-dimensional Weighted and Unweighted Phase Unwrapping that uses Fast Transforms and Iterative Methods", *J. Opt. Soc. Am. A* **11**, 107-117 (1994).
8. D. C. Ghiglia, G. A. Mastin, and L. A. Romero, "Cellular Automata Method for Phase Unwrapping", *J. Opt. Soc. Am. A* **35**, 267-280 (1987).

CHAPTER 7

Conclusions and Further Work

The work in this thesis suggests that the phase distribution associated with an optical disturbance can be successfully recovered from accurate measurements of the intensity information in the image and Fourier domains of an optical system. Consequently, the corresponding optical wave-front can be reconstructed and it will be possible to recreate in a digital computer the discrete complex functions which describe the disturbance in both the spatial and spectral domains.

It has been shown that the phase information plays an important role in the image formation and image reconstruction processes. Phase is either unobservable directly or can not be determined anywhere as accurately as the intensity. On the other hand, magnitude can always be obtained by taking the square root of the measured intensity. Without any additional information or constraints, magnitude and phase are generally independent functions in the sense that the knowledge of one is not sufficient to uniquely deduce the other, and both the magnitude and phase distributions are required in order to reconstruct the corresponding optical wave-front function.

This phase retrieval problem has been successfully solved using an iterative algorithm, based on the Gerchberg and Saxton algorithm, which recovers the phase distribution of an object disturbance from the magnitude information taken from the image and Fourier diffraction domains.

Computer simulations proved that the process of phase retrieval is efficient and converges to the correct solution when the true magnitude information for both the considered domains is used. It has been found that the algorithm converges to incorrect solutions when magnitude information which has been previously calculated from the intensity measurement is used. The ambiguities and inconsistency in the obtained solutions are related to the loss of vital information which is contained in the high frequency components in the Fourier diffraction pattern distribution.

True distributions of the magnitude will never be obtained in practice, so these results suggested that care must be taken to accurately measure and record these quantities from

a real experiment. However, special attention should be given to the Fourier domain where vital information is easily lost due to the low intensity levels of the light usually involved.

The computer simulations have been extended to real experiments. An experimental apparatus has been implemented in order to measure and record on a digital computer the intensity information from the corresponding image and Fourier domains of the optical system. Since the magnitude information calculated from the intensity is a problem in the phase retrieval process, a particular capability of the image sensors on the experimental arrangement has been used to overcome this problem. This capability concerns an electrical operation of gamma correction performed by the sensors which enables the measurement of the quasi-magnitude information instead of intensity due to the increased output signal of the sensors for low intensities of light. The experimental results obtained with the information taken from the experimental apparatus using this gamma correction capability proved that the phase retrieval process converges to the correct solution due to the consistency and reproducibility of the results.

The simulation and experimentation have shown that phase distributions can be successfully recovered which have a valuable and increasingly important role to play in wave-front reconstruction of optical disturbances. The system presented and described has limitations, in particular, it can only be used to recover small phase change distributions over a relatively small aperture, but in principle it would work for any kind and amount of phase change if the scale conditions between the image and Fourier domains can be satisfied.

It has also been shown that discrete complex wave-front functions associated with an optical disturbance can be recreated on the computer using the recovered phase distribution to describe the object disturbance in both the spatial and spectral domains. The ability to perform Fourier transformations of such functions makes possible the generation of schlieren images of the object disturbance on the computer screen. This is a new technique that can be used as a complement to the conventional schlieren method due to the very high sensitivity which has been achieved. The generation of schlieren images on the computer screen has been shown those that could not be obtained using conventional methods due to the low phase gradients for the corresponding sensitivity of the instrument. This technique has also a valuable and important role to play in quantitative schlieren measurements. These measurements normally require several schlieren images of the same event which is often in practice very difficult to reproduce. This technique makes it possible and easy because the event is only performed once, to

acquire and record the necessary information in both the image and Fourier domains, before the use of the algorithm for the phase retrieval process.

Further work should be carried out to investigate the performance and behaviour of the iterative algorithm when more complex object disturbances are used. Work should concentrate first in computer simulations using object modelling and later, if positive results are obtained, an extension of these results to real experiments should be carried out. Scope for further work also exists in improving the phase retrieval algorithm and the computer generation of schlieren images in terms of the interface with the user and processing speed. An automatic scheme to perform physical measurements from the recovered phase distribution of an object disturbance should also be made available bearing in mind that phase contains most of the important information about the disturbance.

PUBLICATIONS

Articles and Conference Papers

The purpose of this last chapter is to give a detailed list of the articles which have been published by the author and the list of the conference papers that have also been presented by the author over the period of study at Loughborough University of Technology.

JOURNALS

- M. T. Carnell, P. T. Fiadeiro and D. C. Emmony, "Cavitation Phenomena in a Lithotripter Shock Field", *Journal of the Acoustic Society of America* **97**, no. 1, 677-679 (1995).
- P. T. Fiadeiro and D. C. Emmony, "Phase Retrieval in Computer Generation of Schlieren Images", *SPIE Proceedings* **2546**, 331-340 (1995).
- P. T. Fiadeiro and D. C. Emmony, "Discrete Reconstruction of Real Phase Objects: a Comparison with Computer-Simulated Phase Objects", *Applied Optics: Information Processing* **34**, 7460-7467 (1995).

CONFERENCES

- The Institute of Physics IOP Annual Congress: Conference Progress in Physical Acoustics and Ultrasonics at Telford (United Kingdom), March 1995, a poster entitled "*Pressure Information Retrieval from Schlieren Images*" has been presented.
- The International Society for Optical Engineering SPIE's 40th Annual Meeting: Conference Optical Techniques in Fluid, Thermal, and Combustion Flow at San

Diego (California USA), July 1995, a paper entitled "*Phase Retrieval in Computer Generation of Schlieren Images*" has been presented.

