UNIVERSITY OF CALGARY

Phase Recovery from Holographic Interferometry Imagery

by

Derek D'Arcy Lichti

A DISSERTATION

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF GEOMATICS ENGINEERING

CALGARY, ALBERTA

FEBRUARY, 1999

National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisitions et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

0-612-38487-X

Canada

ABSTRACT

Holographic interferometry (HI) is an extremely high-resolution optical metrology technique frequently used for deformation measurement and vibration studies. The physical basis of a HI system is the interference of coherent radiation. The phase difference between two wavefronts, one modulated by a vibrating object, is captured as an interference or fringe pattern. The displacement field is implicitly contained in the fringe image as the phase differences. Thus, to obtain the displacement field, the phase difference map must be recovered in some manner. Various methods for phase map recovery have been developed. One such method involves indirect recovery from the fringe intensity image. For the second group of methods, special geometric imaging configurations and signal processing algorithms are exploited for direct phase map recovery.

Existing methods for indirect phase recovery are based upon a search of the fringe image for local intensity maxima that delineate 2π radian contour lines of the underlying phase map. However, such algorithms do not perform well in the presence of noise and non-linear fringe contrast variation. A new algorithm developed for this dissertation exploits the topological relationships between the intensity image fringes and their maxima. Specialised filtering techniques, morphological image processing and computer vision algorithms are utilised to complete the task.

The methods of direct phase recovery, while providing greater resolution displacement field maps, suffer from a discontinuity problem. The numerical processing involved in the recovery has the effect of wrapping the recovered phase. That is,

continuous phase maps are rendered discontinuous on the range $[-\pi,\pi]$ with 2π discontinuities. The task of phase unwrapping is to remove these discontinuities and thereby obtain the continuous phase map. Existing methods treat the problem of two-dimensional unwrapping and a series of independent, one-dimensional unwrapping tasks. A new approach presented in this dissertation utilises topological definitions of phase maps, multi-scale edge detection and computer vision techniques to perform the unwrapping. Also, by approaching the problem as a two-dimensional task, the final step of the algorithm, the unwrapping, is in fact collapsed to a one-dimensional problem.

ACKNOWLEDGEMENTS

During the course of this research, many individuals and organisations provided much needed financial, technical and moral support. The Natural Sciences and Engineering Research Council, The University of Calgary, The Department of Geomatics Engineering and my academic advisor, Dr. M. A. Chapman, are gratefully acknowledged for providing financial support. R. Motriuk, of Trans-Canada Pipelines, Ltd. is gratefully acknowledged for providing the holographic image prints. T. Schmidt, of Holographics. Inc., is thanked for insightful technical conversations. Doctors D. A. Clausi, K. -P. Schwarz, W. F. Teskey, D. Irvine-Halliday and M. A. Chapman are thanked for lending their invaluable time and assistance. Thanks also to colleagues B. Chaplin, P. Tate and D. Provins for many useful technical discussions. Also, thanks are expressed to the Geomatics Engineering secretarial and technical personnel for their assistance. Thanks to my supervisor for challenging and encouraging me throughout my graduate studies. Finally, thanks to my dear wife, Veronica, for her unconditional support throughout this endeavour.

TABLE OF CONTENTS

Approval Page ii
Abstract iii
Acknowledgements v
Table of Contents vi
List of Tables ix
List of Figures x

CHAPTER 1: INTRODUCTION 1
1.1 Deformation Measurement of Industrial Machinery 1
1.1.1 Contact Methods 2
1.1.2 Optical Methods 3
1.1.3 Holography 4
1.2 Research Objectives and Contributions 6
1.3 Dissertation Organisation 8

CHAPTER 2: HOLOGRAPHIC IMAGING 11
2.1 Electromagnetic Waves 11
2.2 Holographic Imaging 15
2.2.1 General Description of a Holographic Imaging System 16
2.2.2 The Incident Radiation 18
2.2.3 Interference Image Formation 20
2.2.4 Interference Image Recording 26
2.2.5 On-Axis Holography 27
2.2.6 Holographic Interferometry 28
2.3 Holographic Image Examples 28
2.4 Error Sources in Holographic Imaging 32
2.4.1 Laser Speckle 32
2.4.2 Film Grain Noise 35
2.4.3 Digitising Noise Sources 36
2.4.4 Non-linear Photographic Recording 37
2.5 Noise Reduction 37
2.6 Chapter Summary 39

CHAPTER 3: PHASE RECOVERY FROM HOLOGRAPHIC IMAGERY 40
3.1 Deformation Measurement from On-Axis Holographic Imagery 40
3.2 Phase Recovery by Phase Stepping 45
3.3 Phase Recovery by the Fourier Transform Method 47
3.4 Phase Unwrapping 53
3.4.1 One-dimensional Phase Unwrapping 56
3.4.2 Two-dimensional Phase Unwrapping 57
3.5 General Case of Holographic Deformation Measurements 59
3.6 Chapter Summary 61

CHAPTER 4: FRINGE MAXIMA EXTRACTION 62
4.1 Theoretical Basis 62
4.2 Description of the New Method 64
4.2.1 Homomorphic Filtering 65
4.2.2 Thresholding 66
4.2.3 Thinning 67
4.2.4 Contour Location and Following 69
4.2.5 Contour Linking and Spur Removal 69
4.2.6 Artefact Identification and Rectification 71
4.2.6.1 Linear Contour Interpolation 71
4.2.6.2 Concentric Contour Matching 72
4.2.6.3 Quadratic Contour Interpolation 74
4.3 Chapter Summary 76

CHAPTER 5: CONTOUR-BASED TWO-DIMENSIONAL PHASE UNWRAPPING 77
5.1 Theoretical Basis of the Contour-Based Method 77
5.2 Description of the Contour-Based Method 81
5.2.1 Edge Detection 82
5.2.2 Edge Location and Following 85
5.2.3 Edge Linking 87
5.2.4 Region Identification 87
5.2.5 Contour-wise Unwrapping 88
5.3 Chapter Summary 91

CHAPTER 6: TESTING AND RESULTS 92
6.1 Fringe Extraction 92
6.1.1 Data Description 92
6.1.2 Pre-filtering 95
6.1.3 Homomorphic Filtering 98
6.1.4 Thresholding 100
6.1.5 Thinning 102
6.1.6 Automatic and Semiautomatic Editing 105
6.1.7 Efficiency and Accuracy of Fringe Maxima Extraction 107
6.1.8 Quantification of Noise Sources 109
6.1.8.1 Speckle 110
6.1.8.2 Film Grain Density 110
6.1.8.3 Sampling 113
6.1.8.4 Quantization 115
6.1.8.5 Other Noise Sources 116
6.1.9 Accuracy Quantification 116
6.1.10 Simulated Image Contour Extraction 118

6.2 Phase Unwrapping 119
6.2.1 Data Description 119
6.2.2 Data Processing 120

CHAPTER 7: SUMMARY, CONCLUSIONS AND RECOMMENDATIONS 123
7.1 Summary 123
7.2 Conclusions 125
7.2 Recommendations for Future Development 128

REFERENCES 130

APPENDIX A: RELEVANT IMAGE TRANSFORMS AND THEOREMS 140
A.1 The Two-dimensional Continuous Fourier Transform 141
A.2 Two-dimensional Continuous Fourier Transform Theorems 142
A.3 The Two-dimensional Discrete-Space Fourier Transform 143
A.4 The Two-dimensional Discrete Fourier Transform 144
A.5 The Fast Fourier Transform 145
A.6 Two-dimensional Discrete Fourier Transform Theorems 146
A.7 Multi-resolution Representation and the Wavelet Transform 147

LIST OF TABLES

Table 6.1. Fringe Maxima Extraction Efficiency 108
Table 6.2. Standard Deviation in Displacement for Different Values of Density 113
Table 6.3. Empirical Fringe Positional Accuracy Statistics 118

LIST OF FIGURES

Figure 2.1. An Off-Axis Holographic Imaging System 16
Figure 2.2. Reference Wave Front Phase Shift Due to Beam Deflection Angle 19
Figure 2.3. Quantized Gaussian Phase Map Surface 29
Figure 2.4. Cross-Section AA' of the Quantized Gaussian Phase Map Surface 29
Figure 2.5. Quantized On-Axis Holographic Image 30
Figure 2.6. Cross-Section AA' of the Quantized On-Axis Hologram 30
Figure 2.7. Quantized Off-Axis Holographic Image 31
Figure 2.8. Cross-Section AA' of the Quantized Off-Axis Hologram 32
Figure 3.1. Fringe Maxima Contours of the On-Axis Hologram of Figure 2.5 (Contrast Reversed) 42
Figure 3.2. Optical Path Length Difference in On-Axis Holographic Interferometry 43
Figure 3.3. Log Magnitude Spectrum of the Off-Axis Hologram of Figure 2.7 49
Figure 3.4. Log Magnitude Spectrum of the Off-Axis Hologram of Figure 2.5 50
Figure 3.5. Wrapped Phase Map 54
Figure 3.6. Cross-Section AA' of the Wrapped Phase Map 54
Figure 3.7. One-dimensional Phase Offset Function 55
Figure 3.8. One-dimensional Unwrapped Phase 56
Figure 3.9. General Case of Optical Path Length Difference Geometry in Holographic Interferometry 59
Figure 4.1. Spur and Bridge Contours 70
Figure 4.2. X-Junction Artefact Resolution 71
Figure 4.3. Concentric Contour Generation Procedure 73
Figure 4.4. Inappropriate Contour Match 74
Figure 4.5. Parabolic Contour Interpolation 75
Figure 4.6. Fringe Maxima Contour Extraction Algorithm 76
Figure 5.1. Regions and Contour Types 78
Figure 5.2. Regions and Contours of the Wrapped Phase Map 79
Figure 5.3. Neighbourhood Search in Edge Following 86
Figure 5.4. Determination of the Relative Phase Offset between Two Regions 89
Figure 5.5. Two-dimensional Phase Unwrapping Algorithm 91
Figure 6.1. Working Images for the Fringe Extraction Algorithm Testing 94
Figure 6.2. Intensity Cross-section along the Central Column of Image 1445 Before Filtering 95
Figure 6.3. Log Magnitude Spectrum of Image 1445 96
Figure 6.4. Symmetric, Gaussian Band-Stop Filters 97
Figure 6.5. Intensity Cross-section along the Central Column of Image 1445 After Filtering 97
Figure 6.6. Homomorphically Filtered Working Images 99
Figure 6.7. Histograms of Image 384 101
Figure 6.8. Global and Local Thresholding Results (Image 1445_2) 102
Figure 6.9. Thinned Fringe Maxima 103

Figure 6.10. Thinned Fringe Maxima of Image 1445 without Homomorphic Filtering 104
Figure 6.11. Edited Fringe Maxima Contours 106
Figure 6.12. Contours Extracted from the Simulated Image of Figure 2.5 119
Figure 6.13. Original and Noisy Simulated Phase Maps 120
Figure 6.14. Edge Detection Magnitude Responses 121
Figure 6.15. Noisy and Original Unwrapped Phase Maps 122
Figure A.1. One-dimensional First Derivative of a Gaussian Wavelets 150

CHAPTER 1

INTRODUCTION

1.1 Deformation Measurement of Industrial Machinery

A problem of great concern in industries such as power generation and oil and gas transmission is that of deformations in the components of heavy rotating machinery such as turbines and generators as well as sections of pipe. Temporally varying deformations induce excessive vibrations in the machinery that can trigger an automatic shut-off system. The resulting unsafe working conditions, reduced productivity and excessive component wear during the vibration as well as lost revenue after shut off are naturally of great concern to plant managers. An industrial rotating machine is a very complex mechanical system, making the diagnosis of the cause of the deformations difficult. It is, therefore, expedient to first measure the deformations with great precision. Only then can trends be analysed and possible causes be hypothesised and solutions formulated.

The precise measurement of deformations of various rotating machinery components necessitates highly specialised methods due to various operational constraints. These constraints are chiefly due to environmental conditions, but are also imposed by the nature of the deformations. First, the industrial setting in which most rotating machinery is operated is often very hazardous and, as a result, remote, non-

contact means of measurement must be utilised. Second, the components of interest, such as shaft bearing housings, are generally situated in relatively inaccessible locations. Third, the magnitude of deformation measurement sought in such diagnostic problems is very often at the sub-millimetre or micrometre levels. Complicating matters further is the fact that the rapidly oscillating vibration is superimposed onto the underlying deformation, which makes rapid data acquisition critical in order to capture the desired information. Ultimately, the temporally varying vibration displacement must be removed from the measurements to obtain the underlying deformation. This can be done with a precisely timed system. Thus, the requirements of the chosen method can be summarised as:

- Non-contact.
- Rapid acquisition time.
- High resolution (micrometre level).
- Provide a dense sampling within a small volume to capture complex deformation fields.

1.1.1 Contact Methods

Contact methods of measuring vibration-induced deformation utilise some form of sensor to measure acceleration, velocity or displacement. Mounted directly on the machinery, an accelerometer senses the force imparted by the motion on a piezoelectric mass. An electric current that is proportional to the force and, hence, acceleration, is

associated with the mass. In order to determine displacement, the acceleration signal must be integrated twice with respect to time. Similarly, a velocity meter signal must be integrated once to obtain a displacement.

While an accelerometer can yield high-resolution displacement measurements, one drawback is that it only provides information at a single location, not over an entire surface. This may necessitate the use of several time-synchronised instruments. Notwithstanding this, accelerometers are often used in conjunction with holographic methods to provide complimentary information for diagnosing problems with heavy machinery.

1.1.2 Optical Methods

Optical methods for the measurement of such deformations include traditional surveying methods as well as photogrammetric techniques. In the former, modern electronic total station instruments are used to survey a network of targeted points affixed to the machinery. By collecting direction, zenith-angle and distance observations, the target positions are estimated via least-squares estimation. The positional accuracy of this approach is on the order of a few hundredths of a millimetre. Bayly (1991) offers an in-depth treatment of surveying methodology in machinery alignment monitoring. An obvious drawback to this approach is the labour-intensive and time-consuming data capture process. Thus, only long-period deformation phenomena can be captured by this means, not higher frequency variations.

Photogrammetric measurement can also be employed for machinery monitoring. By acquiring several high-resolution photographic images of the machinery with a short exposure time, three-dimensional positions of targeted points can be estimated. The accuracy is comparable to that of the previously described surveying methods. For an example of photogrammetric machinery measurement, see Fraser (1985). While photogrammetric data acquisition is much more rapid, it suffers from the requirement of at least two images acquired from different locations and orientations. Photographic images only contain two-dimensional magnitude information since they are exposures of incoherent radiation. Thus, for three-dimensional measurement, two or more images are necessary.

1.1.3 Holography

Holographic interferometry (HI) is an optical method capable of fulfilling all of the requirements listed in Section 1.1. In contrast to the surveying and photogrammetric methods, HI is an active optical method utilising a coherent radiation source (a laser). Furthermore, HI measurement resolution, while varying with laser wavelength, is on the order of a few hundred nanometres, two orders of magnitude greater than the other two methods.

HI has been successfully employed in a variety of structural and mechanical applications. The most common application is deformation measurements (Judge *et al.*, 1992; Pascal *et al.*, 1996; Pouet and Krishnaswamy, 1996; and Quan *et al.*, 1996). Other uses include vibration analysis (Pryputniewicz, 1989; Beeck, 1992), structural inspection

of aircraft (Schmidt and Webster, 1996) and biomedical measurement (Pryputniewicz, 1989).

The concept of holographic imaging was first published by Denis Gabor in 1948, but was not physically realised until the invention of the laser in the 1960s. The basis of holographic imaging is the interference of two coherent wavefronts, one of them modulated in phase by the object being measured, such as the bearing housing of a machine. The product of this system is a recorded fringe pattern of alternating contrast. Areas with dark tone correspond to destructive interference between the two wave fronts, while light tones indicate constructive interference. Contained within the fringe pattern image is the phase modulation, or phase difference surface, as a function of two-dimensional position. It is this phase surface, which is proportional to the third dimension (i.e., depth) of the object, that gives holography its most unique, and powerful, property: a two-dimensional holographic image contains three-dimensional information. This is a stark contrast to a photographic image, which contains only two-dimensional information. HI is a simple extension of holography in which a second exposure of the same film is made at a later time. The resulting phase modulation characterises the change in shape (deformation) of the object between exposures.

In order to recover the deformation information about an object, the phase differences must be recovered from the holographic recording. This can be accomplished through a series of image processing operations to extract the fringe maxima from the digitised interferogram. The algorithm documented by Robinson and Reid (1993) performs this task by searching the image for local intensity maxima. However, the

success of this technique has been limited to noise-free imagery with high contrast. A new approach to this problem described in this dissertation utilises several image processing and computer vision techniques and exploits topological rules of fringe imagery.

For the greatest resolution, the phase difference surface is extracted via a sequence of filtering operations. However, due to the numerical nature of the recovery process, only the principal value of the phase surface is obtained. That is, the recovered phase difference surface is mapped from its full amplitude onto the range $(-\pi, \pi)$ by the modulus operation and, as such, is discontinuous. The crux of holographic image processing is the unwrapping of this wrapped phase surface, which amounts to removal of the discontinuities. The solutions presented thus far, varied as they may be, all have a common thread in that the unwrapping of a two-dimensional phase surface is treated as a series of independent, one-dimensional unwrapping problems. A new approach to this problem is presented in this dissertation, and is briefly outlined in the next section.

1.2 Research Objectives and Contributions

The objectives of this research effort are to develop, implement and test improved algorithms for both intensity-based interferogram analysis and two-dimensional phase unwrapping. The existing means of deformation extraction from fringe patterns utilises operators for the detection of local intensity maxima (Robinson and Reid, 1993). However, as will be shown in Chapter Six, the noise characteristics of the HI imagery used for this research do not permit the use of this approach. The new approach

presented here first makes use of homomorphic filtering to reduce the multiplicative speckle noise. Fringe maxima are then extracted via local thresholding. Subsequent thinning of the maxima via morphological processing yields single pixel wide contours. These contours are then tracked using computer vision techniques. Various algorithms for bridging gaps in the contours complete the process.

As previously mentioned, the two-dimensional unwrapping problem has been approached by other researchers as a sequence of independent, one-dimensional problems. The shortcoming of this approach is that the strong correlation between neighbouring phase map samples is not exploited. The benefit of utilising this correlation is that spurious discontinuities as well as gaps in the true discontinuity contours can be eliminated and bridged, respectively. Such defects arise from the various noise sources and distortions inherent in digitised holographic imagery.

The new approach to this problem that is described herein treats the problem as a fully two-dimensional problem. Andrä *et al.* (1991) first tackled two-dimensional phase unwrapping in this manner, utilising various computer vision techniques employing edge detection and line following. The algorithm developed for this dissertation builds upon this idea, utilising scale space concepts.

The contributions of this research to the field of holographic interferometry are given below.

- *A topological description of both fringe imagery and two-dimensional phase maps.* In order to devise the aforementioned algorithms, it was first necessary to formulate the basic spatial relationships between the various entities, such as the fringe maxima

of an interferogram and the discontinuity contours and intermediary regions of a phase map.

- *The design, implementation and testing of an improved algorithm for fringe maxima extraction.* The new algorithm uses many different image-processing techniques to overcome the difficulties posed by speckle noise and non-linear modulation depth in fringe maxima extraction.
- *The design, implementation and testing of a two-dimensional unwrapping algorithm using computer vision techniques multiple scale filtering.* Like Andrä *et al.* (1991), the new algorithm utilises various image segmentation techniques such as edge detection, edge following, edge linking and region growing. However, the new contribution is in the combination of edge detector responses from different scales (window widths).
- *An algorithm that reduces a two-dimensional unwrapping problem to a one-dimensional unwrapping problem.* It will be shown that, by adopting a two-dimensional topological approach to phase unwrapping, the actual unwrapping is reduced to a one-dimensional problem.

1.3 Dissertation Organisation

Chapter Two contains the fundamentals of holographic image formation. The chapter begins with the mathematical description of electromagnetic waves with emphasis on coherent radiation sources (lasers) and their properties. Holographic imaging systems are then addressed, followed by the mathematical derivation of the

holographic image formation process and detailed analysis of the resulting equation. Both on-axis and off-axis holographic images are studied. Finally, the various noise sources, both systematic and random, which distort holographic images, are addressed.

Existing methods of phase recovery for both on-axis and off-axis holography are described and compared in Chapter Three. Existing methods of phase unwrapping follow this section. Finally, the derivation of deformations from the recovered phase is also presented and a detailed error analysis is given.

In Chapter Four, the algorithm for fringe maxima extraction is presented. First, the topology of fringe imagery is described. The filtering operations are then detailed, followed by the thresholding algorithm. Next, morphological processing and thinning are discussed. The chapter concludes with the discussion of the various gap-bridging strategies.

The new two-dimensional phase unwrapping algorithm developed for this dissertation is described in Chapter Five. First, a topological description of wrapped phase imagery is outlined. Second the chronological steps of the algorithm are detailed. Third, the advantages of the new algorithm over existing approaches are discussed.

Both simulated and real holographic imagery was used to test the new algorithm's performance. Chapter Six begins with a detailed description of the parameters of these data sets, followed by test results and data analysis. Finally, a summary of this research, and recommendations for future work are given in Chapter Seven.

The subject of Appendix A is two-dimensional transforms. Integral transforms, and their discrete counterparts, play significant roles in the phase recovery and fringe

maxima extraction processes, and must, therefore, be included in this dissertation. The two-dimensional continuous Fourier transform is first presented, along with some related theorems that are drawn upon in this manuscript. The discrete version of this transform, along with its fast computational algorithm, is then treated. Finally, the chapter concludes with scale space filtering and the wavelet transform, which are utilised in the new unwrapping algorithm.

CHAPTER 2

HOLOGRAPHIC IMAGING

The study of holographic imaging must begin with a description of the nature and analytical form of coherent radiation. This chapter commences with the relevant physical background of monochromatic waves. Based upon this material, the process of holographic image formation is then derived. Both on-axis and off-axis holography are treated. A discussion of both systematic and random error sources logically follows.

2.1 Electromagnetic Waves

Electromagnetic (EM) radiation is a vector field quantity comprised of mutually orthogonal electric, E, and magnetic field, B, components. Each component is orthogonal to the direction of propagation. In this dissertation, only the electric component is treated, as it is more significant in photographic recording. Because of the wave nature of EM radiation, it propagates as periodic waves that are functions of time, t, and spatial position, $\underline{r}$. (In all subsequent discussion, adopted co-ordinate frames will be Cartesian). Because of this periodic nature, EM radiation satisfies the wave equation:

$$\nabla^2 E(t,\underline{r}) = \frac{1}{v^2}\frac{\partial^2 E(t,\underline{r})}{\partial t^2}\ , \tag{2.1}$$

where v is the phase velocity of the wave.

EM radiation from most natural and artificial sources possesses a broad frequency spectrum. That is, it consists of the superposition of waves of many different frequencies or wavelengths. A special case is a monochromatic radiation source that consists of a single frequency and propagates as a pure sinusoid. The frequency spectrum of such a wave is an impulse. In reality, purely monochromatic sources do not exist. Laser light, which is highly monochromatic source with a very narrow bandwidth spectrum, is a close approximation to an ideal source.

Monochromatic waves are characterised by a number of parameters. The frequency can be parameterised in terms of circular frequency, ω, in radians per second or by temporal frequency, f, in Hertz (cycles per second). The wavelength, λ, is related to frequency by the phase velocity. Phase velocity is the velocity of light in a vacuum, c, reduced by the index of refraction of the propagation medium, n. The relationship between these quantities is given by

$$\lambda = \frac{2\pi v}{\omega} = \frac{v}{f} = \frac{c}{nf} \quad . \tag{2.2}$$

The position vector, $\underline{r}$, and propagation vector, $\underline{k}$, define the position and direction of propagation of an EM wave, respectively. The norm of $\underline{k}$ is given by

$$\|\underline{k}\| = \frac{2\pi}{\lambda} \quad . \tag{2.3}$$

The normalised propagation vector is a set of direction cosines, which indicate the direction of wave travel.

Since EM radiation is a vector quantity, it has up to three sinusoidal components, each with possibly different amplitude. The electric field amplitude is expressed in units of Newtons per Coulomb or, equivalently, Volts per metre. The polarisation of a wave indicates the directional dependence, if any, of the oscillations of the field components as the wave travels through space. For example, the electric field of a wave linearly polarised in the y-axis direction will only oscillate in a direction parallel to that axis. The propagation direction and the direction of polarisation thus define the plane of vibration. Highly monochromatic sources, such as lasers, are designed to produce highly polarised light, while most other sources produce randomly polarised light.

The electric field equation of a monochromatic wave, linearly polarised in the y-direction, is given by

$$\underline{E}(t,\underline{r}) = \begin{pmatrix} 0 \\ E_y(\underline{r})\cos(\omega t + \underline{k} \bullet \underline{r}) \\ 0 \end{pmatrix} . \qquad (2.4)$$

The amplitude of this wave is given by E_y.

Coherence is the fundamental property of laser radiation upon which holographic imaging is based. Coherence is subdivided into two classes: temporal and spatial. Temporal coherence is a measure of the monochromacity of the radiation. A purely monochromatic source has infinitesimal bandwidth and emits infinite duration sinusoidal waves, all of which are in phase. Real sources have finite bandwidth and emit finite duration wave trains, which manifest themselves as truncated sinusoids. Wave trains are actually a superposition of sinusoids of different frequencies.

A useful temporal coherence metric is the coherence length, L_c, given by

$$L_c = \frac{v}{BW} \quad . \tag{2.5}$$

Due to the inverse relationship between coherence length and bandwidth, BW (in Hertz), sources with narrow bandwidth produce highly coherent light. The coherence length indicates the spatial distance at which two sinusoids of a wave train will be 180° out of phase. Stated differently, coherence length is the average wave train length. The importance of this property will be revisited in subsequent sections dealing with holographic image formation.

The wave in Equation 2.4 can be further identified as a plane wave since it is comprised of planar wave fronts. A wave front is a surface defined by the locus of points of constant phase. Spatial coherence is a measure of degree of planarity of the wave fronts. Distortions in wave front shape are caused by phase differences between wave trains emitted from different positions on the radiation source.

It is worth mentioning an alternative representation of a plane wave, which, in some instances, has distinct advantages. To illustrate this formulation, the y-component of Equation 2.4 can be rewritten as:

$$\begin{aligned} E_y(t,\underline{r}) &= \mathrm{Re}\left\{E_y(\underline{r})e^{j(\omega t+\underline{k}\bullet\underline{r})}\right\} \\ &= \mathrm{Re}\left\{U(\underline{r})e^{j\omega t}\right\} \end{aligned} \tag{2.6}$$

where

$$U(\underline{r}) = E_y(\underline{r})\,e^{j\underline{k}\bullet\underline{r}} \quad . \tag{2.7}$$

Re {•} indicates the real part of a complex quantity. The advantage gained with this notation is that the spatial and temporal components of the wave are separated.

An important derived quantity of EM radiation is irradiance, measured in Watts per square metre. The importance of this quantity stems from the fact that photographic film is sensitive to the average power falling on its surface, not the high frequency variations of the waves themselves. Irradiance is the time average radiant flux (power or energy flow rate) of the radiation per unit area falling on a body. Considering the y-axis component of Equation 2.7, the expression for irradiance in a vacuum is given by

$$I = \varepsilon_o c \left\langle E_y(\underline{r}) E_y^*(\underline{r}) \right\rangle . \tag{2.8}$$

The ε_o term is the dielectric constant, or electrical permitivity, of a vacuum. The dielectric constant is frequency dependent and varies with temperature. The angular brackets denote the time averaging operation of the quantity therein, which will be explicitly defined in the next section. The superscripted asterisk indicates complex conjugation.

2.2 Holographic Imaging

The components of an off-axis holographic system, schematically illustrated in Figure 2.1, include the coherent radiation source, the laser, various optical elements, and the recording medium. The recording medium is typically a glass plate coated with a photographic emulsion, although digital recording is becoming more commonplace. The deforming object to be measured may represent a bearing housing or a section of a pipe.

The general operation of the system will be first explained conceptually then mathematically.

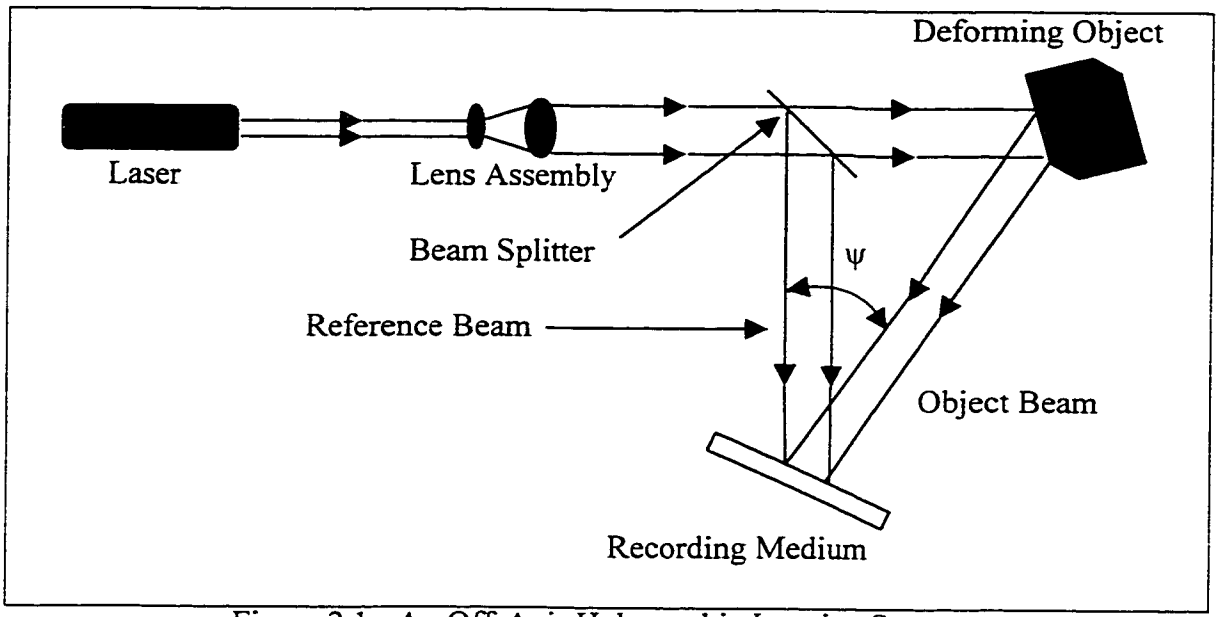


Figure 2.1. An Off-Axis Holographic Imaging System.

2.2.1 General Description of a Holographic Imaging System

The monochromatic beam (comprised of planar wave fronts) emitted from the laser is spread out and collimated back into planar wave fronts, via a lens assembly, in order to increase the system field of view. The beam amplitude is then divided, via the beam splitter, into two components: the *reference* and *object beams*. Note that the beam splitter may also be placed before the lens assembly. The reference beam propagates directly to the recording medium, while the object beam proceeds to the object to be measured. Upon interacting with the object surface, the object beam is diffusely reflected and modulated in both magnitude and phase. The phase modulation represents the difference in optical path length between the object and reference beams. Since the phase

differences are proportional to the object surface topography, they are of most interest in holographic imaging.

Because the two beams are coherent, they interfere when they converge upon the recording medium. The irradiance of the interference pattern is recorded on the film. The phase differences between the two beams are encoded within the interference pattern as a fringe pattern of sinusoidal contrast. Dark tone fringes indicate areas where the two beams are out of phase (destructive interference), while light tone fringes indicate where the beams are in phase (constructive interference). The fringe image thus contains the phase difference between the beams, and hence the topographic structure of the object, as a function of two-dimensional position. Stated differently, three-dimensional information about the object is captured in a two-dimensional image. Reliable extraction of this dimensional information is the goal of holographic image processing.

The distinction between off-axis and on-axis holography is made by noting the presence of the deflection angle, ψ, between the object and reference beams in the off-axis configuration. While the object beam is nominally coincident with the image plane surface normal, the reference beam is tilted by ψ. The purpose of this tilt is to facilitate numerical recovery of the phase differences via spatial frequency domain filtering operations. It will be shown in the next section that the deflection angle is a linear function of position in the image plane. In on-axis or Gabor holography, the deflection angle is zero.

2.2.2 The Incident Radiation

Assuming that the radiation source is linearly polarised coherent light, the electric field component of the reference beam in the plane of the recording medium is given by

$$E_r(x,y) = r(x,y)\cos(\omega t + \theta(x,y)) \ , \tag{2.9}$$

where,

$E_r(x,y)$ is the electric field component of the reference beam in the image plane,

$r(x,y)$ is the electric field amplitude of the reference beam in the image plane,

$\theta(x,y)$ is the phase shift in the reference beam due to the deflection angle ψ.

To derive the expression for $\theta(x,y)$, the deflection angle geometry will first be considered in one dimension, then extended to two dimensions. Referring to Figure 2.2, the deflection angle ψ introduces the geometric shift in phase, Δ, into the reference beam wave front as it strikes the image plane. From the geometry of Figure 2.2,

$$\Delta = x\sin\psi \ . \tag{2.10}$$

To obtain the phase shift in angular units (radians), Equation 2.10 is simply scaled and subsequently simplified as

$$\frac{2\pi}{\lambda}\Delta = \frac{2\pi}{\lambda}x\sin\psi$$

$$\theta(x) = \left(\frac{2\pi}{\lambda}\sin\psi\right)x = \omega_o x \tag{2.11}$$

Thus, the phase shift due to the constant deflection angle ψ is a linear function of spatial position in the image plane. Extending this development to two dimensions, the expression for $\theta(x,y)$ is given by

$$\theta(x,y) = \omega_o x + \nu_o y \ . \tag{2.12}$$

The spatial frequencies ω_o and ν_o are called carrier frequencies and are given in radians per linear unit, such as millimetres. Their significance in the phase recovery process will be explored in Chapter Three.

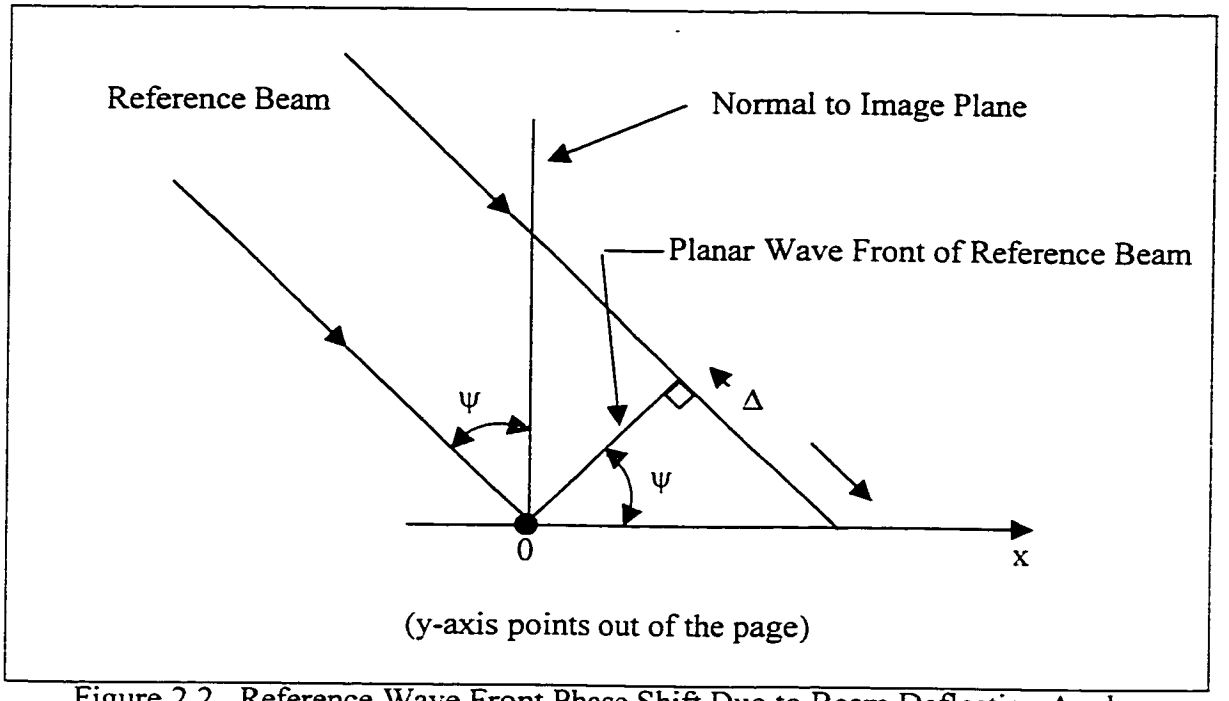


Figure 2.2. Reference Wave Front Phase Shift Due to Beam Deflection Angle.

The electric field component of the object beam in the plane of the recording medium is given by

$$E_s(x,y) = s(x,y)\cos(\omega t + \phi(x,y)), \tag{2.13}$$

where,

$E_s(x,y)$ is the electric field component of the object beam in the image plane,

$s(x,y)$ is the electric field amplitude of the object beam in the image plane,

$\phi(x,y)$ is the phase shift in the object beam, also called the phase map or phase distribution.

Note that the reference and object beam amplitudes differ due to the modulation. Furthermore, the phase map caused by the object surface is generally non-linear.

2.2.3 Interference Image Formation

Due to the coherence of the radiation, the superposition of these two beams is simply the addition of the two electric field components

$$E(x,y) = E_r(x,y) + E_s(x,y) \ . \tag{2.14}$$

The irradiance of the interfering waves is determined using Equation 2.8.

$$\begin{aligned} I(x,y) &= \varepsilon_o c\left\langle E(x,y)E^*(x,y)\right\rangle \\ &= \varepsilon_o c\left\langle (E_r(x,y)+E_s(x,y))(E_r(x,y)+E_s(x,y))^*\right\rangle \\ &= \varepsilon_o c\left\langle E_r(x,y)^2 + E_s(x,y)^2 + 2E_r(x,y)E_s(x,y)\right\rangle \\ &= \varepsilon_o c\left\{\left\langle E_r(x,y)^2\right\rangle + \left\langle E_s(x,y)^2\right\rangle + 2\left\langle E_r(x,y)E_s(x,y)\right\rangle\right\} \end{aligned} \tag{2.15}$$

Note that irradiance of the superposition of two incoherent beams is calculated as the sum of the irradiances of the individual waves. As will be shown shortly, it is the coherence property of laser radiation that permits the existence of the third term in Equation 2.15. Without this term, as would be the case with incoherent imaging, no fringe pattern would be visible.

Following the procedure of Pedrotti and Pedrotti (1993), the analytical form of the irradiance is derived term by term. The time average operator for a sinusoidal waveform, with period T, may be defined as

$$\langle f(t)\rangle = \frac{1}{T}\int_0^T f(t)\,dt \ . \tag{2.16}$$

Using this definition and Equation 2.9, the first term of Equation 2.15 becomes

$$\left\langle E_r(x,y)^2\right\rangle = \frac{1}{T}\int_0^T r(x,y)^2\cos^2(\omega t+\theta(x,y))\,dt \ .$$

Using the identity

$$\cos^2\alpha = \frac{1}{2}+\frac{1}{2}\cos 2\alpha \tag{2.17}$$

and extracting the temporally independent term from the integrands gives

$$\left\langle E_r(x,y)^2\right\rangle = \frac{r(x,y)^2}{2T}\left\{\int_0^T dt+\int_0^T \cos 2(\omega t+\theta(x,y))\,dt\right\} .$$

Using the identity

$$\cos(\alpha+\beta) = \cos\alpha\cos\beta-\sin\alpha\sin\beta \tag{2.18}$$

and again extracting the temporally independent terms gives

$$\left\langle E_r(x,y)^2\right\rangle = \frac{r(x,y)^2}{2T}\left\{\int_0^T dt+\cos 2\theta(x,y)\int_0^T \cos 2\omega t\,dt\right.$$

$$\left.-\sin 2\theta(x,y)\int_0^T \sin 2\omega t\,dt\right\}$$

The first integral is equal to T, and the last two are zero as they are the time averages of pure sinusoids. Thus, the final expression for the first term of Equation 2.15 is

$$\left\langle E_r(x,y)^2 \right\rangle = \frac{r(x,y)^2}{2} \quad . \tag{2.19}$$

Equation 2.19 indicates that the irradiance of the reference beam is proportional to one half of square of the electric field amplitude. Through identical development, it can be shown that the second term of Equation 2.15 is.

$$\left\langle E_s(x,y)^2 \right\rangle = \frac{s(x,y)^2}{2} \quad . \tag{2.20}$$

This equation represents the irradiance of the object beam.

Using Equations 2.9, 2.13 and 2.16, the third term of Equation 2.15 is

$$2\langle E_r(x,y)E_s(x,y)\rangle = \frac{2}{T}\int_0^T r(x,y)s(x,y)\cos(\omega t + \theta(x,y))\bullet$$

$$\cos(\omega t + \phi(x,y))dt$$

Using Equation 2.18 and performing some algebraic manipulation, this expression becomes

$$\begin{aligned} 2\langle E_r(x,y)E_s(x,y)\rangle = \frac{2r(x,y)s(x,y)}{T}\bullet & \\ \Big\{ \cos\theta(x,y)\cos\phi(x,y)\int_0^T \cos^2\omega t\,dt & \\ -\cos\theta(x,y)\sin\phi(x,y)\int_0^T \sin\omega t\cos\omega t\,dt & \\ -\sin\theta(x,y)\cos\phi(x,y)\int_0^T \sin\omega t\cos\omega t\,dt & \\ +\sin\theta(x,y)\sin\phi(x,y)\int_0^T \sin^2\omega t\,dt \Big\} & \end{aligned}$$

Using the identity

$$\sin(\alpha+\beta)=\sin\alpha\cos\beta+\cos\alpha\sin\beta\ , \tag{2.21}$$

the above expression becomes

$$\begin{aligned}2\langle E_r(x,y)E_s(x,y)\rangle = \frac{2r(x,y)s(x,y)}{T}\bullet \\ \Big\{\cos\theta(x,y)\cos\phi(x,y)\int_0^T\cos^2\omega t\,dt \\ -\sin(\theta(x,y)+\phi(x,y))\int_0^T\sin\omega t\cos\omega t\,dt \\ +\sin\theta(x,y)\sin\phi(x,y)\int_0^T\sin^2\omega t\,dt\Big\}\end{aligned}$$

Since

$$\langle\cos^2\omega t\rangle=\langle\sin^2\omega t\rangle=\frac{1}{2}\ , \tag{2.22}$$

and

$$\langle\sin\omega t\cos\omega t\rangle=0\ , \tag{2.23}$$

this equation reduces to

$$\begin{aligned}2\langle E_r(x,y)E_s(x,y)\rangle = r(x,y)s(x,y)\bullet \\ \{\cos\theta(x,y)\cos\phi(x,y)+\sin\theta(x,y)\sin\phi(x,y)\}\end{aligned}$$

Finally, using the following identity

$$\cos(\alpha-\beta)=\cos\alpha\cos\beta+\sin\alpha\sin\beta\ , \tag{2.24}$$

the third term of Equation 2.15 becomes

$$2\langle E_r(x,y)E_s(x,y)\rangle = r(x,y)s(x,y)\cos(\theta(x,y)-\phi(x,y))\ . \tag{2.25}$$

Substitution of Equations 2.19, 2.20 and 2.25 into Equation 2.15, the expression for hologram irradiance is

$$I(x,y)=\varepsilon_o c\left\{\frac{r(x,y)^2+s(x,y)^2}{2}\right. \tag{2.26}$$

$$\left.+r(x,y)s(x,y)\cos(\theta(x,y)-\phi(x,y))\right\}$$

Equation 2.26 can also be expressed in terms of the irradiance of both the reference and object beams as:

$$I(x,y)=\left\{\frac{I_r(x,y)+I_s(x,y)}{2}\right.$$

$$\left.+\sqrt{I_r(x,y)I_s(x,y)}\cos(\theta(x,y)-\phi(x,y))\right\} \tag{2.27}$$

$$=\frac{I_r(x,y)+I_s(x,y)}{2}\{1+M(x,y)\cos(\theta(x,y)-\phi(x,y))\}$$

where

$I_r(x,y)$ is the irradiance of the reference beam in the image plane:

$$I_r(x,y)=\varepsilon_0 c\, r(x,y)^2 \ , \tag{2.28}$$

$I_s(x,y)$ is the irradiance of the object beam in the image plane:

$$I_s(x,y)=\varepsilon_0 c s(x,y)^2 \ , \tag{2.29}$$

$M(x,y)$ is the modulation depth:

$$M(x,y)=\frac{2\sqrt{I_r(x,y)I_s(x,y)}}{I_r(x,y)+I_s(x,y)} \ . \tag{2.30}$$

As a final step, Equation 2.12 is substituted into Equation 2.27, giving

$$I(x,y)=\frac{I_r(x,y)+I_s(x,y)}{2}\{1+M(x,y)\cos(\omega_o x+\nu_o y-\phi(x,y))\}\ . \qquad (2.31)$$

Analysis of Equation 2.31 reveals its structure to be a constant term plus an amplitude-modulated sinusoidal term. In reality, the dc term, or average irradiance, is not truly constant throughout the image plane, but is a low frequency phenomenon. The importance of this behaviour will be demonstrated in the subsequent discussion on phase recovery.

The cosine term contains the difference between the linear phase shifts, induced by the off-axis geometric configuration, and the desired phase map. The maximum value of this term is one, which occurs when its argument is zero radians. Physically, this represents perfect phase alignment between the reference and object beams and corresponds to constructive interference (bright fringes). When the argument is equal to $\pm\pi$, radians, the cosine minimum of negative one is realised. This situation occurs when the two beams are out of phase by one half of one cycle and produces destructive interference (dark fringes).

Preceding the cosine term is the modulation depth, or fringe visibility (Caulfield and Lu, 1970), which determines the degree of contrast between light and dark fringes. If the modulation depth is allowed to approach zero, contrast is low and the holographic image is dominated by the average irradiance. In this scenario, the phase map is difficult -- if not impossible -- to recover. Thus, maximum fringe contrast is desirable. In general,

the modulation depth is a non-linear function, but varies slowly throughout the image plane.

Modulation depth is affected by the geometric configuration of the holographic imaging system. If the difference in optical path length between the reference and object beams is greater than the coherence length, no fringes will be visible. Thus, the path lengths must be carefully matched. Equally as important, though, is a laser with a sufficiently long coherence length to fulfil the measurement task. Collectively, the product of the modulation depth and cosine terms is called the interference term. This term disappears when the radiation source is incoherent and diminishes when the coherence length is exceeded by the optical path length difference.

2.2.4 Interference Image Recording

As previously mentioned, the photographic film used to make holographic recordings is sensitive to irradiance. Exposure, H, is defined as the product of irradiance and the exposure time, Δt, and has units of Joules per square metre (Caulfield and Lu, 1970).

$$H(x,y) = I(x,y)\Delta t \tag{2.32}$$

Using this definition, the exposure of the hologram irradiance given by Equation 2.31 is

$$H(x,y) = \frac{I_r(x,y)+I_s(x,y)}{2}\{1 + M(x,y)\cos(\omega_o x + \nu_o y - \phi(x,y))\}\Delta t$$

$$= H_o + \left\{\sqrt{I_r(x,y)I_s(x,y)}\cos(\omega_o x + \nu_o y - \phi(x,y))\right\}\Delta t \tag{2.33}$$

where

H_o is the average exposure.

Transmittance is the ratio of transmitted flux to incident flux. If the hologram is recorded in the linear portion of the film transfer function (H-T curve), then the resulting transmittance is of the form

$$T(x,y) = A + BH(x,y) , \tag{2.34}$$

where A and B are constants. The form of the transmittance, or intensity, image is given by Equation 2.35.

$$T(x,y) = i(x,y) = a(x,y) + b(x,y)\cos(\omega_o x + \nu_o y - \phi(x,y)) . \tag{2.35}$$

where

$i(x,y)$ is the recorded image intensity,

$a(x,y)$ is the background intensity,

$b(x,y)$ is the fringe visibility.

Note that Equation 2.35 is proportional to the irradiance (Equation 2.31). The problem faced in holographic image processing is to extract the phase map from digital version of Equation 2.35.

2.2.5 On-Axis Holography

As will be described in the chapter concerned with phase recovery, the off-axis geometry (see Figure 2.2) is employed as a means to facilitate precise, and possibly automated, phase map recovery. In on-axis holography, the deflection angle ψ is equal to

zero. Later in this chapter, the difference between the two types of holographic imagery, on-axis and off-axis, will be illustrated.

2.2.6 Holographic Interferometry

A common task of holographic imaging is the measurement of changes in object shape (deformations) rather than its actual shape. This can be accomplished by utilising the fact that holographic recording is a linear process (in complex amplitude) in the sense that two images can be superimposed (Vest, 1979). That is, if a holographic image of an object is recorded onto film at time t_1, and then a second image is recorded on the same film at time t_2, the resulting fringe pattern will contain the phase map proportional to the change in object shape. The intensity image will be of the same form as Equation 2.35. However, the linear, spatial phase shifts are introduced between the two exposure times. If no phase shifts are introduced, then the resulting HI image is effectively an on-axis interferogram.

2.3 Holographic Interference Image Examples

In order to illustrate the fundamental differences between on-axis and off-axis holography, some simulated images are presented. The first step was generation of a two-dimensional phase map representing the phase differences induced by a fictitious surface. The phase surface chosen was a circularly symmetric Gaussian function ranging from zero to 12π radians. This surface could represent the first bending mode of a thin, uniformly loaded plate clamped at each corner. Although the Gaussian function may not

be the true analytical shape of the bending mode, it is a simple function to generate and is thus useful for illustrative purposes. Figure 2.3 is an eight-bit quantized representation of this phase map, in which black and white represent the minimum and maximum phase map amplitudes, respectively. While the symmetry is evident in this image, the Gaussian shape clearly reveals itself in the intensity cross section AA', shown in Figure 2.4.

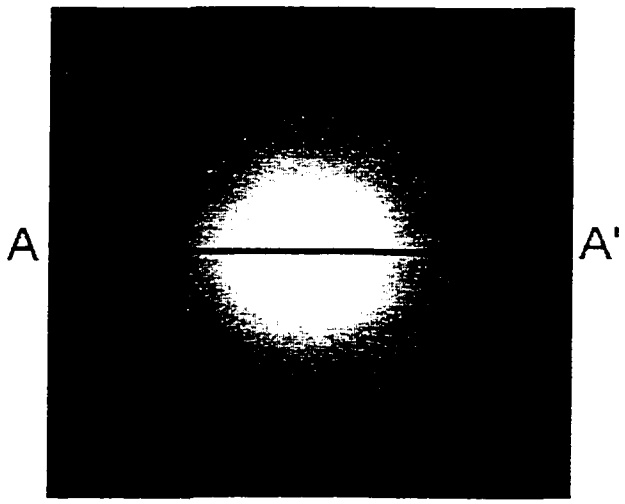


Figure 2.3. Quantized Gaussian Phase Map Surface.

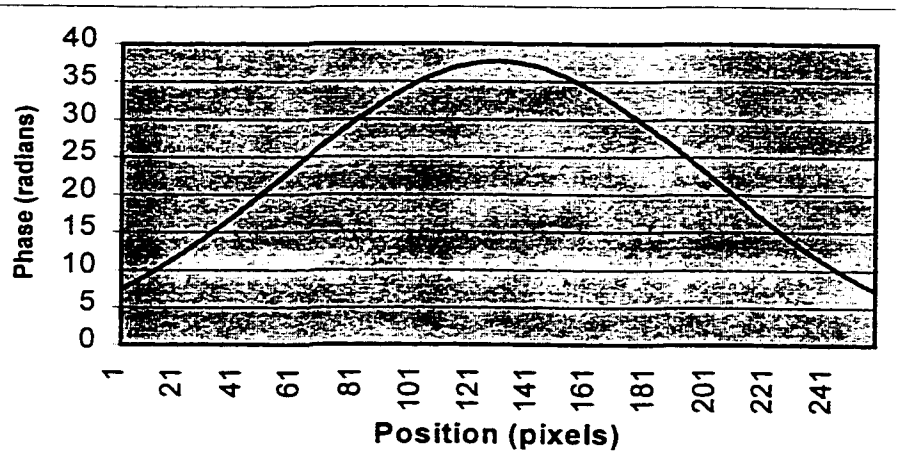


Figure 2.4. Cross-Section AA'of the Quantized Gaussain Phase Map Surface.

Substituting the generated phase map into Equation 2.35, and setting the other required parameters, holographic images were then generated. The on-axis hologram, for which $\omega_o \nu_o = 0$, is shown in Figure 2.5 and cross section AA′ is shown in Figure 2.6.

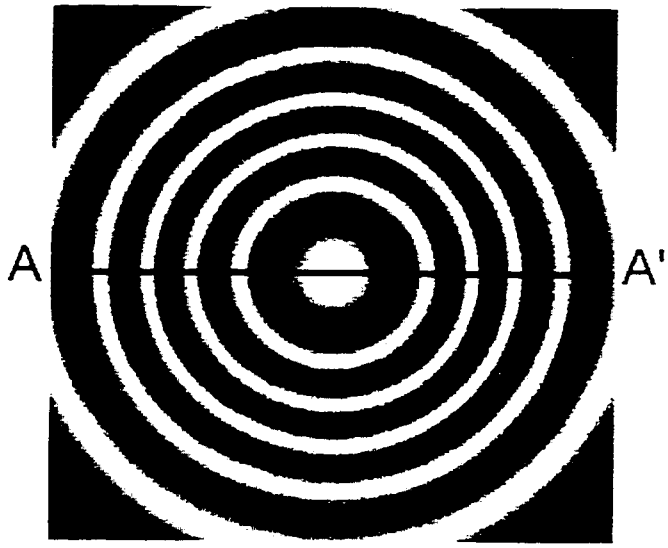


Figure 2.5. Quantized On-Axis Holographic Image.

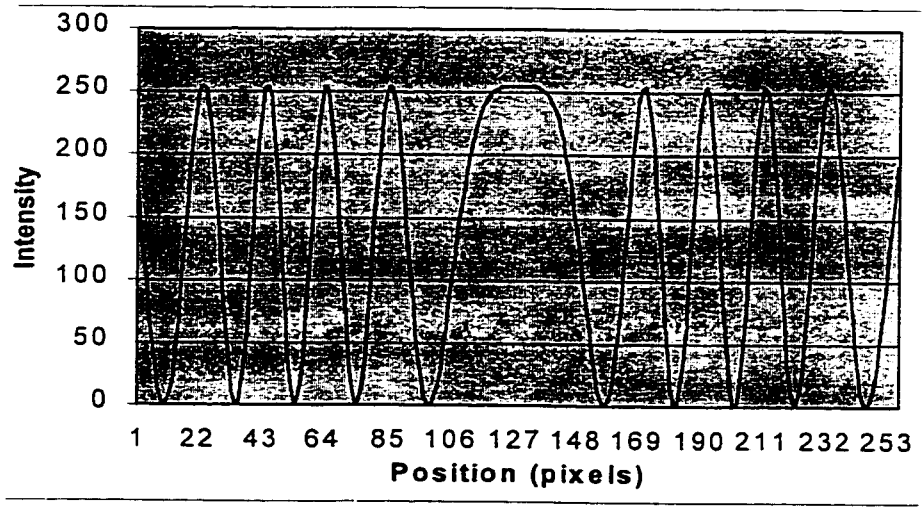


Figure 2.6. Cross-Section AA′of the Quantized On-Axis Hologram.

The sinusoidal intensity variation is clearly visible in both the image and cross section figures. Note that the symmetry of the underlying phase map is preserved by the on-axis geometry.

The off-axis holographic image resulting from the introduction of non-zero spatial carrier frequencies is shown in quantized intensity in Figure 2.7 and in cross section in Figure 2.8. The carrier frequencies have the effect of increasing the spatial frequency of the fringe pattern, as can be seen in both figures. Also note that the symmetry of the phase map is lost in the hologram image formation.

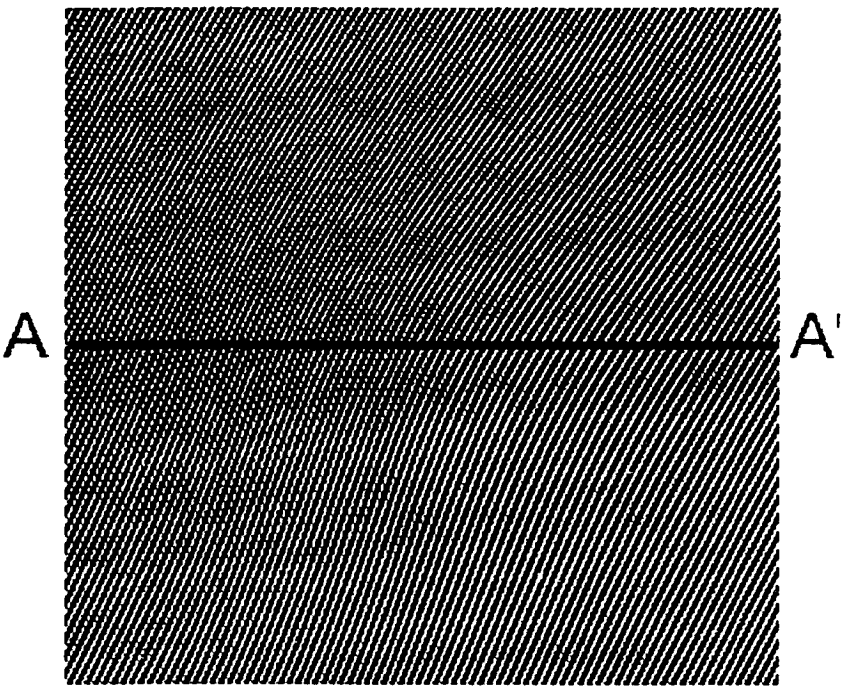


Figure 2.7. Quantized Off-Axis Holographic Image.

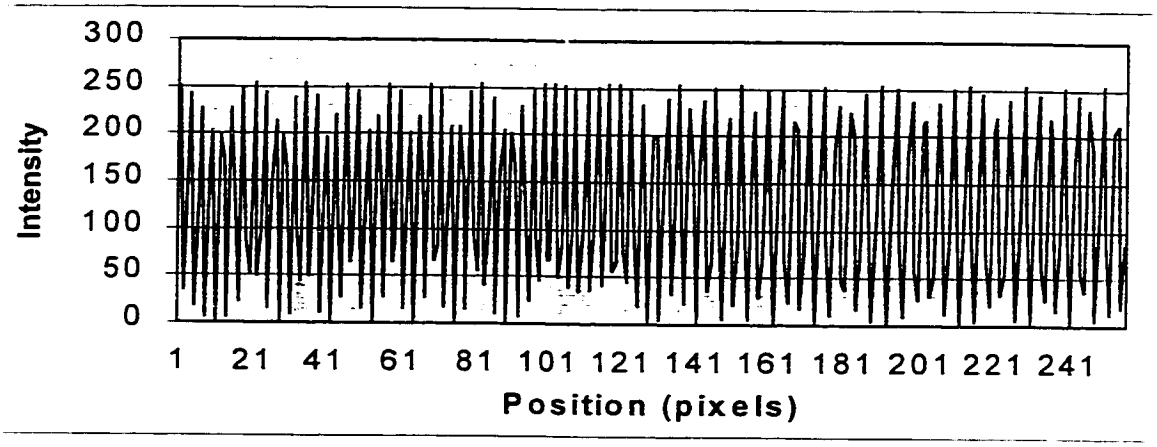


Figure 2.8. Cross-Section AA'of the Quantized Off-Axis Hologram.

2.4 Error Sources in Holographic Imaging

As with any method of metrology, there are various errors inherent in the acquired data. These include both systematic and random errors. The distinguishing feature of holographic imaging is that both additive and multiplicative random errors exist. Additive errors are due to non-uniform film grain density of the photographic emulsion, as well as electronic and quantization noises introduced by the image digitising process. Multiplicative noise, known as speckle, arises due to the coherent radiation source as well as the surface roughness of the object. Non-linear photographic recording is a systematic error that is treated at the end of this section.

2.4.1 Laser Speckle

Speckle noise is a random phenomenon encountered in coherent imaging methods such as holography and synthetic aperture radar. The cause of this effect is the highly

coherent nature of the radiation source. Random variations in object surface topography on the order of the laser wavelength cause random interference patterns in the recorded holographic image. This gives the hologram a speckled appearance.

The basis of the stochastic model of speckle is the fact that, due to the diffuse reflectance, the speckle pattern at any point in the hologram will contain contributions from many points on the object surface. That is, in contrast to pinhole photographic image model, there is no one-to-one correspondence between object points and image points. Thus, the speckle interference pattern is a summation of random complex amplitude contributions from many different points on the object surface, as shown in Equation 2.36 (Goodman, 1974):

$$A(\underline{r}) = \sum_{k=1}^{N} |a_k| e^{j\theta_k}, \tag{2.36}$$

where

$A(\underline{r})$ is the complex amplitude due to the speckle pattern,

$|a_k|$ is the amplitude of random contribution k,

θ_k is the phase of random contribution k,

N is the number of contributions, which is assumed to be large.

The resulting model is that of a random walk in the complex plane.

Two critical assumptions about the stochastic properties of speckle are the amplitude and phase are independent, and the phase is uniformly distributed on $[-\pi, \pi]$ (Goodman, 1976). Goodman (1984) proved that the amplitude is normally distributed.

Of greatest interest, though, is the density of the speckle irradiance, p(I), which Goodman (1984) shows to be negative exponential

$$p(I)=\frac{1}{\mu_I}e^{\frac{-I}{\mu_I}}u(I)\ , \tag{2.37}$$

where

μ_I is the mean value of the irradiance,

$u(I)$ is the unit step function.

Analysis of Equation 2.37 reveals that the speckle irradiance can not assume negative values. Furthermore, the most probable value of speckle irradiance is zero. An interesting property of this density is the standard deviation, σ_I, is equal to the mean, μ_I (Vest, 1979)

Also of interest in the stochastic description of a speckle irradiance is its spatial correlation. The auto-correlation function gives a measure of the average width of the speckle pattern. As shown by Vest (1979), the spatial correlation is a function of laser frequency, object surface geometry and imaging geometry. For the case of a rectangular surface of dimensions $L \times L$, wavelength λ, and an object-image plane separation of z, the auto-correlation function is a raised sinc-squared function (Goodman, 1984).

$$R_I(\Delta x,\Delta y)=\mu_I^2\left\{1+\mathrm{sinc}^2\left(\frac{L\Delta x}{\lambda z}\right)\mathrm{sinc}^2\left(\frac{L\Delta y}{\lambda z}\right)\right\}, \tag{2.38}$$

where

$R_I(\Delta x,\Delta y)$ is the auto-correlation function of speckle intensity,

$\Delta x,\Delta y$ are the spatial separations within the image plane in x and y, respectively,

and

$$\mathrm{sinc}(x) = \frac{\sin \pi x}{\pi x} . \tag{2.39}$$

The average width of the speckle pattern, δx, is defined as the distance to the first zero of the sinc-squared function in Equation 2.37, which is (Goodman, 1984)

$$\delta x = \frac{\lambda z}{L} . \tag{2.40}$$

As mentioned at the outset of this section, laser speckle is treated as a multiplicative noise source. The model of hologram image intensity degraded by multiplicative noise is given by (Lim and Nawab, 1981)

$$i'(x,y) = w(x,y)i(x,y) , \tag{2.41}$$

where

$i(x,y)$ is the holographic image intensity, as given by Equation 2.35.

$w(x,y)$ is the multiplicative speckle noise.

$i'(x,y)$ is the degraded hologram.

2.4.2 Film Grain Noise

Film grain noise arises from non-uniform concentration or distribution of the film grains in the photographic emulsion (Lee, 1972). This has the effect of introducing random variation in the background intensity of an image. A common model for film grain noise is white, Gaussian additive noise (Castleman, 1996). Addition of this effect to Equation 2.41 yields

$$i'(x,y) = w(x,y)i(x,y) + n(x,y)\ , \tag{2.42}$$

where

$n(x,y)$ is the additive film grain noise.

2.4.3 Digitising Noise Sources

Other important noise sources in holographic imagery are due to the digitising process. In order to obtain a numerical representation of an image, it is necessary to use a scanning device that utilises electro-optical sensors to convert intensity into voltage and, ultimately, digital numbers. Noise is added at several junctures in the process. The additive noise sources in this section can be combined with film grain noise term $n(x,y)$ of Equation 2.41.

Electronic noise arises due to random electron motion in the digitising circuit elements and is modelled as a zero-mean, white Gaussian process (Castleman, 1996). Photo-electronic noise is due to random variations in the incident illumination required for scanning and can be considered as a Gaussian process at high illumination levels (Castleman, 1996). Once the voltage signal representation of the image is obtained from the acquisition hardware, it is uniformly sampled and quantized into digital numbers. Quantization is a non-linear process in which the signal amplitude is either rounded or truncated to integer values to minimise computer storage requirements. Treated as an additive noise source, quantization error has a uniform density (Oppenheim and Schafer, 1989).

2.4.4 Non-linear Photographic Recording

In Subsection 2.2.3, the transformation from exposure to transmittance was presumed to take place on the linear portion of the H-T curve. In reality, the H-T transfer function is non-linear, and exposure must be carefully regulated so that the near-linear portion of the curve is utilised.

Caulfield and Lu (1970) investigated the effects of recording in the non-linear portion of the H-T curve in great detail. The analysis is conducted by performing a Taylor series expansion of transmittance, Equation 2.35, about the average exposure, H_o, as shown in Equation 2.39.

$$T = T_o + \left.\frac{dT}{dH}\right|_{H=H_o} (H - H_o) + \frac{1}{2!}\left.\frac{d^2T}{dH^2}\right|_{H=H_o} (H - H_o)^2 + \frac{1}{3!}\left.\frac{d^3T}{dH^3}\right|_{H=H_o} (H - H_o)^3 + \ldots \tag{2.43}$$

Caulfield and Lu (1970) showed that the second order and higher terms of the Taylor series contribute to the background intensity of the hologram and as well as harmonic sinusoidal terms.

2.5 Noise Reduction

As a result of the various error sources outlined in the previous section, a holographic image is corrupted by both additive and multiplicative noise. Broadband, additive noise can be reduced by low-pass filtering. Two of the simplest approaches for dealing with the multiplicative noise are low-pass filtering and homomorphic filtering. In

the former approach, the speckle noise is assumed to be spatially uncorrelated and thus a broadband phenomenon (Lim and Nawab, 1981). This assumption permits the decomposition of both the intensity and noise terms of Equation 2.41 into their low (L) and high (H) frequency components

$$\begin{aligned} i'(x,y) &= (w_L(x,y) + w_H(x,y)) \cdot (i_L(x,y) + i_H(x,y)) \\ &= i_L(x,y)w_L(x,y) + i_L(x,y)w_H(x,y) \\ &\quad + i_H(x,y)w_L(x,y) + i_H(x,y)w_H(x,y) \end{aligned} \quad . \tag{2.44}$$

Since most of the image content is concentrated at low frequencies, the last two terms of Equation 2.44 contribute little to the degraded image intensity. By low-pass filtering the image, the high frequency component of the noise (the second term of Equation 2.44) is greatly attenuated. Neglecting the last two terms, the remaining filtered image is approximately

$$i'_{LPF}(x,y) \cong i_L(x,y)w_L(x,y) \; . \tag{2.45}$$

Thus, by low-pass filtering the intensity, the signal-to-noise ration has been improved as most of the speckle noise has been removed while most of the image content has been retained (Lim and Nawab, 1981). Robinson and Reid (1993) also suggest this approach to speckle noise reduction.

Alternatively, homomorphic filtering can be used in which Equation 2.41 is transformed logarithmically, which yields two additive terms.

$$\ln\{i'(x,y)\} = \ln\{w(x,y)\} + \ln\{i(x,y)\} \; . \tag{2.46}$$

Since most of the image content of the density (logarithm of intensity) is at low frequencies and the logarithm of the noise is a broadband field, low-pass filtering Equation 2.46 thus improves the signal-to-noise ratio (Lim and Nawab, 1981). The improved intensity image is recovered by the taking exponential of the filtered density image. This method, as well as the low-pass filtering technique, indeed reduces speckle noise, but does so at the expense of image sharpness due to the blurring property of low-pass filters.

2.6 Chapter Summary

The goal of this chapter was to mathematically describe the physical process of holographic image formation. This has provided the necessary background material for the solving the underlying problem in this thesis: the extraction of a phase distribution, and thus a deformation field, from HI imagery. The error sources inhibiting the success of this task have also been described in detail.

CHAPTER 3

PHASE RECOVERY FROM HOLOGRAPHIC IMAGERY

The ultimate goal of holographic image processing is to extract dimensional information about an object, such as deformation measurements, from the recorded hologram. This task boils down to extraction of the phase difference map from the imagery. Both the manner in which this is accomplished and the available resolution of the recovered phase varies with the geometric configuration of the holographic imaging system. Phase map and deformation recovery from on-axis holography is described first. Following this, the phase stepping technique, which can be visualised as a hybrid of on-axis and off-axis holography, is described. The so-called Fourier transform technique used for phase recovery from off-axis holographic images is then detailed. The final step of phase recovery is called phase unwrapping. Existing methods for unwrapping the phase are reviewed in this chapter. A new, improved method is detailed in Chapter Five.

3.1 Deformation Measurements from On-Axis Holographic Imagery

Recall that in on-axis holographic imaging, the propagation vectors of the object and reference beams are coincident, and the deflection angle, ψ (see Figure 2.2), is equal

to zero. As a result, no spatial carrier frequencies are introduced into the intensity equation, thus reducing Equation 2.35 to

$$\begin{aligned} i(x,y) &= a(x,y) + b(x,y)\cos(-\phi(x,y)) \\ &= a(x,y) + b(x,y)\cos(\phi(x,y)) \end{aligned} \tag{3.1}$$

The second form of Equation 3.1 results from the even symmetry of the cosine function. The equivalence of these two forms poses a fundamental problem in the actual phase recovery, as will be described shortly.

With reference to the on-axis holographic image and corresponding intensity cross section (see Figures 2.5 and 2.6, respectively), each successive fringe maximum (white fringe peak) represents a change in the phase map of 2π radians. Thus, the first step of the phase recovery process entails delineation of the intensity maxima within the image. The fringe maxima contours are shown in Figure 3.1. The next step is to assign a fringe order number, N, to each region lying between the contours. This establishes the topological relationship between each pair of adjacent regions, i.e., which region is higher.

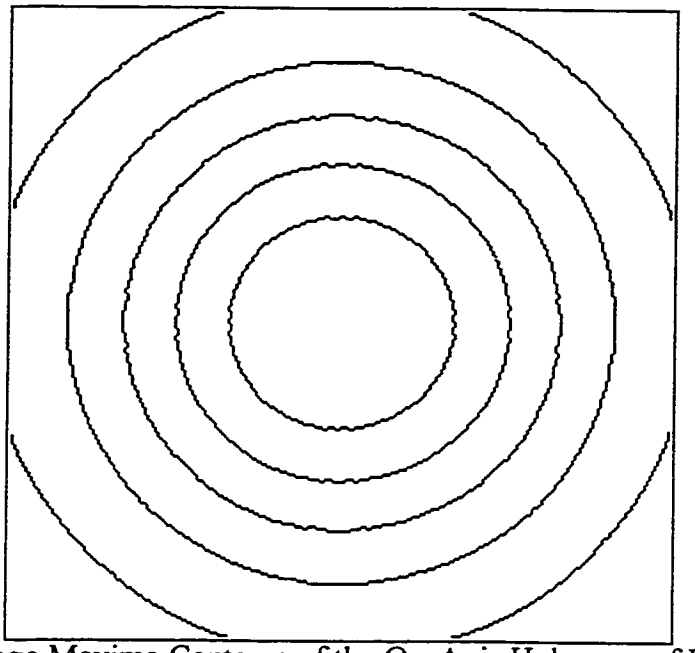


Figure 3.1. Fringe Maxima Contours of the On-Axis Hologram of Figure 2.5 (Contrast Reversed).

The next step is to establish the geometric relationship between the map of phase differences and the quantities of interest, the deformations. Although full three-dimensional information about an object, or its deformation, is contained within a holographic interferogram, it is often only the normal or out-of-plane component that its desired. Recovery of all three components, however, requires detailed knowledge about the geometric position of the hologram relative to the object and the propagation vectors. This case will not be treated here, but details can be found in Vest (1979) and Pryputniewicz (1989).

Figure 3.2 illustrates the case of an object undergoing normal deformation where the propagation vector of the object beam is normal to the object surface. When the object is in its original state, the object wave front propagates along the optical path CBC.

When the object is in its deformed state, the object wave front propagates along the path described by CAC. The optical path length difference BAB is proportional to the phase difference between the two beams.

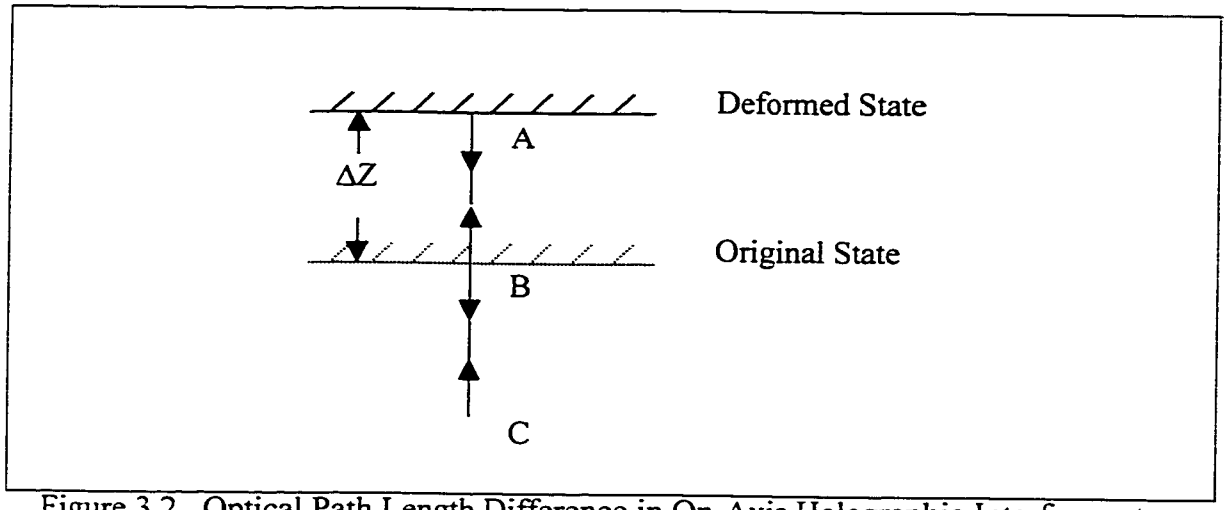


Figure 3.2. Optical Path Length Difference in On-Axis Holographic Interferometry.

Once the fringe order numbers have been assigned, they are converted to out-of-plane deformations through the optical path length differences. The path length difference BAB, denoted as $\Delta\ell$, is linearly proportional to the phase difference

$$\Delta\ell(x,y)=\frac{\lambda}{2\pi}\phi(x,y)\ . \tag{3.2}$$

The phase difference, ϕ, is resolved in terms of the fringe order numbers, N, which are integer multiples of 2π radians

$$\begin{aligned}\Delta\ell(x,y)&=\frac{\lambda}{2\pi}2\pi N_i\\ &=\lambda N_i\end{aligned}\ , \tag{3.3}$$

where N_i is the order number of the i^{th} fringe. The normal component of the object deformation, ΔZ, is equal to, for the geometry of Figure 3.2, half the optical path length difference

$$\Delta Z(x,y) = \frac{\Delta \ell(x,y)}{2}$$

$$= \frac{\lambda}{2} N_i \tag{3.4}$$

For on-axis holographic interferometry, the out-of plane deformation at any point in the image is simply equal to the fringe order number multiplied by one half of the laser wavelength. Although this represents a measurement of extremely high resolution, there are shortcomings. First, because the phase is only resolved to integer multiples of 2π radians, fractional phase differences can not be resolved, even though they are captured in the fringe pattern. Second, the recovered phase pattern and, hence, deformation surface, is inherently ambiguous. Due to the equivalence of the two forms of Equation 3.1, it is impossible to analytically determine if a recovered surface is concave upward or downward without a priori knowledge of the deformation experienced by the object. Thus, intervention is required to make this decision. Direct methods of phase recovery, detailed next, are free of this ambiguity, but possess a different ambiguity problem. However, if only peak deflection is required as in vibration analysis, the concavity ambiguity problem is not an issue.

One solution for fringe maxima extraction from HI imagery, as documented by Robinson and Reid (1993), utilises a bank of local operators for peak detection. Fringe peaks are marked if the image gradient in two orthogonal directions meet certain criteria.

As pointed out by Robinson and Reid (1993), this method works well in high-contrast, noise-free imagery. A new method that tackles the fringe maxima extraction problem in noisy imagery with significant contrast variation is proposed in Chapter Four.

3.2 Phase Recovery by Phase Stepping

In the phase stepping method, a known phase shift is introduced into the reference beam rather than the deflection angle, ψ (see Figure 2.2). Phase stepping is described by Kujawinska (1987) for Moiré interferometry, but is equally applicable to holography. For this method of phase recovery, a minimum of three holographic images is required. The phase shifts are usually introduced with a mirror that can be rotated in discrete steps. Common values for the phase shifts are -90°, 0°, 90° and 0°, 120°, 240°.

To illustrate the phase recovery process through phase stepping, assume that the induced phase shifts are -90°, 0° and 90°. The equation corresponding to the i^{th} reference wave front is

$$E_r(x,y) = r(x,y)\cos(\omega t + \alpha_i) , \tag{3.5}$$

where $\alpha_i = i \cdot 90°$ for $i \in \{-1,0,1\}$. Following the same procedure detailed in Chapter Two, the i^{th} intensity image is

$$i(x,y) = a(x,y) + b(x,y)\cos(\alpha_i - \phi(x,y)) . \tag{3.6}$$

Substitution of the values for α_i gives the expressions for the three intensity images

$$\begin{aligned} i_{-1}(x,y) &= a(x,y) + b(x,y)\cos(-90° - \phi(x,y)) \\ &= a(x,y) - b(x,y)\sin(\phi(x,y)) \end{aligned} \tag{3.7a}$$

$$i_0(x,y) = a(x,y) + b(x,y)\cos(\phi(x,y)) \tag{3.7b}$$

$$i_1(x,y) = a(x,y) + b(x,y)\cos(90° - \phi(x,y))$$

$$= a(x,y) + b(x,y)\sin(\phi(x,y)) \tag{3.7c}$$

The difference between Equations 3.7c and 3.7a is

$$i_1(x,y) - i_{-1}(x,y) = 2b(x,y)\sin(\phi(x,y)) \ , \tag{3.8}$$

while the sum of Equations 3.7c and 3.7a is

$$i_1(x,y) + i_{-1}(x,y) = 2a(x,y) \ . \tag{3.9}$$

Two times Equation 3.7b minus Equation 3.9 gives

$$2i_0(x,y) - (i_1(x,y) + i_{-1}(x,y)) = 2b(x,y)\cos(\phi(x,y)) \ . \tag{3.10}$$

The rational expression of Equation 3.8 over Equation 3.10 is

$$\frac{i_1(x,y) - i_{-1}(x,y)}{2i_0(x,y) - (i_1(x,y) + i_{-1}(x,y))} = \frac{2b(x,y)\sin(\phi(x,y))}{2b(x,y)\cos(\phi(x,y))}$$

$$= \tan(\phi(x,y)) \tag{3.11}$$

Finally, the expression for the phase is

$$\phi(x,y) = \arctan\left\{ \frac{i_1(x,y) - i_{-1}(x,y)}{2i_0(x,y) - (i_1(x,y) + i_{-1}(x,y))} \right\} \ . \tag{3.12}$$

Equation 3.12 indicates that the phase difference at any sample location in a holographic image can be recovered as the arctangent of the ratio of linear combinations of intensity. Also note that the phase at any point is recovered independently of its neighbours. The phase recovery is inaccurate when both the numerator and denominator of Equation 3.12 are small (Judge *et al.*, 1992).

The advantage of this method over the on-axis technique is direct recovery of the phase map. However, the success of phase stepping requires that the object remains stationary between each exposure. For monitoring industrial machinery which, in addition to undergoing deformation, is vibrating, three static images may be difficult to acquire.

3.3 Phase Recovery by the Fourier Transform Method

The so-called Fourier transform method is based upon a series of linear filtering operations for which the FFT algorithm is exploited to reduce the computational burden of the phase recovery. This method was first developed by Takeda *et al.* (1982) as a one-dimensional algorithm applied to each row or column of a sampled off-axis holographic image. Bone *et al.* (1986) published a fully two-dimensional application of the method. The key to this method is the deflection angle, ψ (see Figure 2.2), between the reference and object beams. By introducing the spatial carrier frequencies, ω_o and ν_o, some of the spectral components of the holographic image spectrum are separated, making possible phase recovery through linear filtering.

First, the holographic intensity image, given by Equation 2.35, is rewritten in terms of complex exponential terms using the identity

$$\cos\theta = \frac{1}{2}\left(e^{j\theta} + e^{-j\theta}\right) . \tag{3.13}$$

The resulting intensity image becomes

$$i(x,y) = a(x,y) + \frac{b(x,y)}{2}\left\{e^{j(\omega_o x + \nu_o y - \phi(x,y))} + e^{-j(\omega_o x + \nu_o y - \phi(x,y))}\right\} . \tag{3.14}$$

By defining

$$c(x,y) = \frac{b(x,y)}{2} e^{-j\phi(x,y)} , \tag{3.15}$$

and

$$c^*(x,y) = \frac{b(x,y)}{2} e^{j\phi(x,y)} , \tag{3.16}$$

Equation 3.14 becomes

$$i(x,y) = a(x,y) + c(x,y)e^{j(\omega_o x + \nu_o y)} + c^*(x,y)e^{-j(\omega_o x + \nu_o y)} . \tag{3.17}$$

The Fourier transform of Equation 3.17, using the frequency-shifting theorem (Equation A.4) and the complex conjugate function theorem (Equation A.6) is given by

$$I(\omega,\nu) = A(\omega,\nu) + C(\omega - \omega_o, \nu - \nu_o) + C^*(\omega + \omega_o, \nu + \nu_o) . \tag{3.18}$$

The logarithm of the magnitude spectrum of the off-axis hologram of Figure 2.7 is given in Figure 3.3. The frequency origin is located at the centre of the image. Note that there are three distinct lobes in the spectrum. The narrow lobe located at the origin is the spectrum of the background intensity, $A(\omega,\nu)$. The other two are conjugate symmetric lobes containing the spectral content of $C(\omega,\nu)$ and $C^*(\omega,\nu)$, respectively. These lobes are centred at the positive and negative spatial carrier frequencies, (ω_o,ν_o) and $(-\omega_o,-\nu_o)$, respectively.

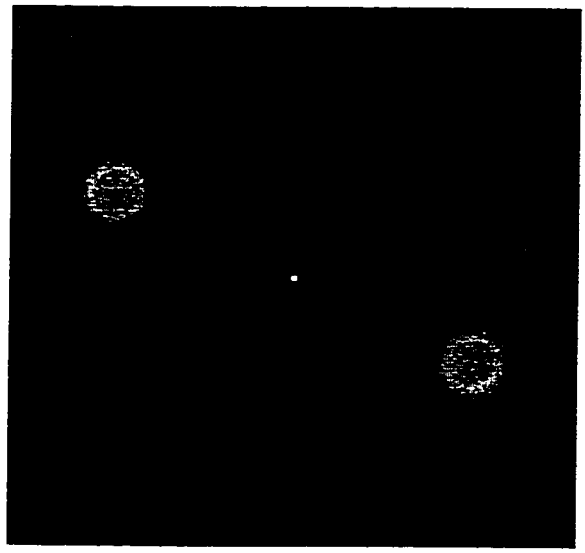


Figure 3.3. Log Magnitude Spectrum of the Off-Axis Hologram of Figure 2.7.

Figure 3.4 shows the log magnitude spectrum of the on-axis hologram of Figure 2.5. In contrast to the off-axis hologram spectrum, the spectrum in Figure 3.4 consists only of a single lobe centred at the frequency origin. All of the image contents, background intensity, fringe visibility and phase map, are contained in this single lobe. Thus, the advantage gained by the introduction of wave front tilt in off-axis holography is spectral separation of the background intensity -- a low-frequency phenomenon -- from the fringe visibility and phase map.

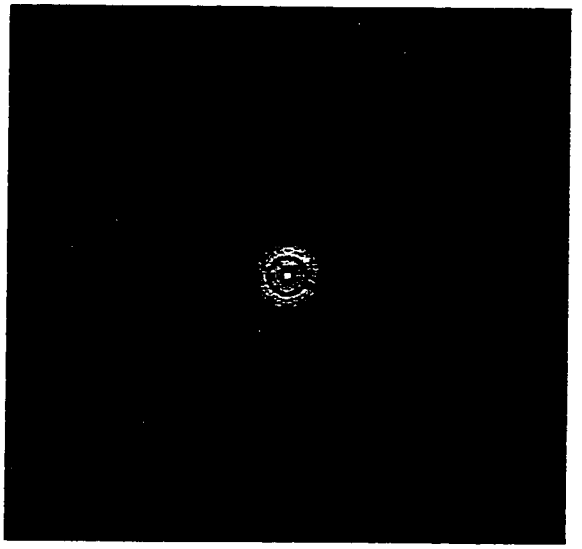


Figure 3.4. Log Magnitude Spectrum of the On-Axis Hologram of Figure 2.5.

The recovery of the phase map from Equation 3.18 is essentially a demodulation problem, which is well known in the field of communication engineering. First the spectrum is band-pass filtered such that one of the two conjugate symmetric lobes are preserved and the other two lobes discarded. If the retained lobe is in the positive-positive frequency quadrant, and assuming ideal filtering, the resulting spectrum is

$$I'(\omega,\nu) = C(\omega-\omega_o,\nu-\nu_o)\ . \tag{3.19}$$

The spatial domain dual of Equation 3.19, given by the inverse Fourier transform, is

$$i'(x,y) = c(x,y)e^{j(\omega_o x+\nu_o y)}\ . \tag{3.20}$$

Next, Equation 3.19 is shifted in the frequency domain by $(-\omega_o,-\nu_o)$. In general the operation of frequency shifting is called heterodyning, while the specific case of shifting

the spectrum to the origin is called base banding. The analytical form of the base banded signal in the frequency domain becomes

$$I''(\omega,\nu)=C(\omega,\nu)\ , \tag{3.21}$$

while in the spatial domain it is

$$\begin{aligned} i''(x,y) &= c(x,y)e^{j(\omega_o x+\nu_o y)}e^{-j(\omega_o x+\nu_o y)} \\ &= c(x,y) \\ &= \frac{b(x,y)}{2}e^{-j\phi(x,y)} \\ &= \frac{b(x,y)}{2}\{\cos(\phi(x,y))-j\sin(\phi(x,y)) \\ &= Re(x,y)-jIm(x,y) \end{aligned} \tag{3.22}$$

The phase map, $\phi(x,y)$, can be recovered from the complex spatial domain image, $i''(x,y)$as

$$\phi(x,y)=\arctan\left\{\frac{-Im(x,y)}{Re(x,y)}\right\}\ . \tag{3.23}$$

The principal advantage of the FFT method is only one holographic image is required for phase recovery. However, this method is more computationally intensive than the phase stepping method, even with the advantages of FFT spectrum computation. Furthermore, distortions due to erroneous frequency shifts, filter bandwidth and non-ideal filter implementation may be present in the recovered phase map.

Erroneous frequency shifting results from inadequate knowledge of the deflection angle, ψ, between the reference and object beams. The result of erroneous frequency shifting is a residual shift, $(\delta\omega, \delta\nu)$, in the spectrum of Equation 3.21 (Bone *et al.*, 1986)

$$I''(\omega,\nu) = C(\omega - \delta\omega, \nu - \delta\nu) \; . \tag{3.24}$$

Using the frequency-shifting theorem, the corresponding spatial domain signal is

$$i''(x,y) = \frac{b(x,y)}{2} e^{-j(\phi(x,y) + \delta\omega x + \delta\nu y)} \; . \tag{3.25}$$

Thus, the effect of incorrect frequency shifts is systematic addition of a linear function to the phase map. In the solution proposed by Bone *et al.* (1986), the induced linear trend is estimated in an area of the phase map that is known to be flat. The resulting plane is then subtracted from the rest of the phase map to remove the effect.

There are two possible effects of inadequate filter bandwidth determination. One possibility is that the bandwidths are too small, in which case the pass band of the spectrum is inappropriately truncated and/or attenuated. The result is loss of phase map information. On the other hand, if the chosen bandwidths are too large, then spectral content belonging to the background intensity may be passed. The result is distortion of the phase map.

Filter selection is an equally important aspect in Fourier transform phase recovery. Since ideal band pass filters introduce ringing into the resulting spatial domain signal, approximations such as Gaussian and low-order Butterworth filters are often used. However, due to the broad transition band nature of these functions, there is monotonic attenuation of the spectral content in the intended pass band.

To circumvent some of these potential distortions, it is advisable to introduce the largest possible tilt (deflection angle ψ) between the object and reference beams. This will maximise the spectral separation of the hologram contents. In this case, a large filter bandwidth can be chosen such that spectral content within the pass band will be less distorted and the content of the background intensity will not be passed.

3.4 Phase Unwrapping

The advantage of the phase stepping and the FFT methods is that the actual phase map is recovered, rather than an inferred map from fringe maxima. The recovered map is not only unambiguous in terms of concavity, but is of greater resolution than that of on-axis holography as fractional phase differences are recovered. The latter point will be demonstrated in the next section. However, these two methods suffer from a different type of ambiguity problem, which inherently stems from the fact that fractional phase differences are recovered.

Recovery of the phase using both the phase stepping and FFT techniques involve the use of the arctangent function. Angles determined numerically via the arctangent can only be determined to modulo 2π radians. Thus, a phase surface obtained using either of these methods is mapped from its true values onto the range $[-\pi, \pi]$ radians. The resulting wrapped phase map has discontinuities every 2π radians.

The image in Figure 3.5 is the wrapped phase map recovered from the off-axis hologram of Figure 2.7. The discontinuous nature of the phase map is revealed by the abrupt tonal transitions from black to white. The cross section shown in Figure 3.6 also reveals the discontinuities.

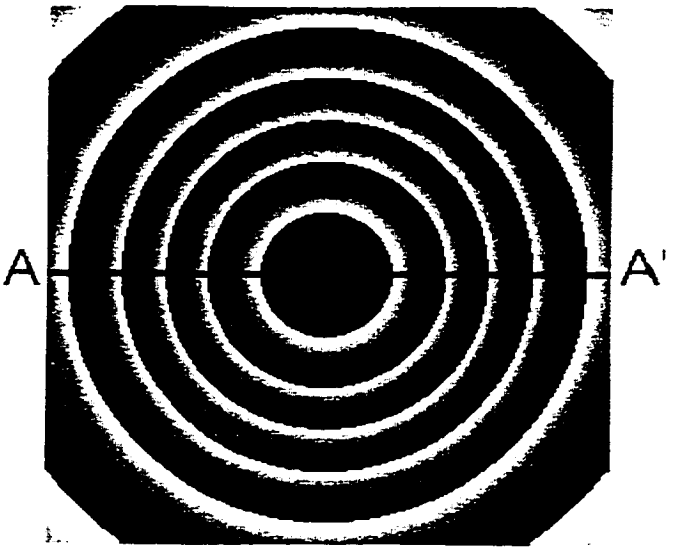


Figure 3.5. Wrapped Phase Map.

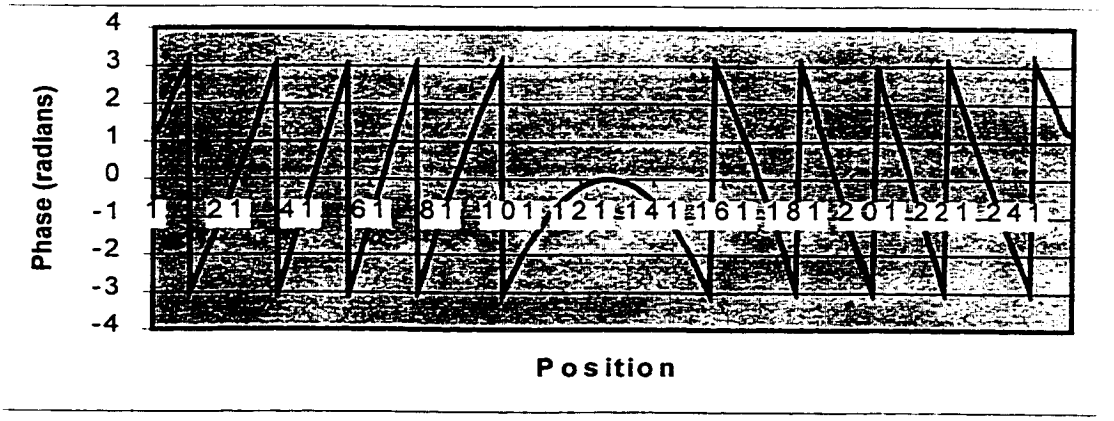


Figure 3.6. Cross-Section AA'of the Wrapped Phase Map.

The objective of phase unwrapping is to remove the discontinuities from the recovered phase map in order to obtain a continuous surface. The problem is one of finding a phase offset function that has values equal to the correct number of integer multiples of 2π radians. The relation between the wrapped, discontinuous phase map, $\phi_d(x,y)$, the phase-offset function $\phi_o(x,y)$ and the continuous phase map $\phi_c(x,y)$ is given by (Takeda *et al.*, 1981).

$$\phi_c(x,y) = \phi_d(x,y) + \phi_o(x,y). \tag{3.26}$$

For the one-dimensional signal shown in Figure 3.6, the phase offset is illustrated in Figure 3.7. Addition of this signal to the wrapped phase yields the continuous phase map shown in Figure 3.8. Note the Gaussian shape of this graph, which resembles that of the original phase map given in Figure 2.3.

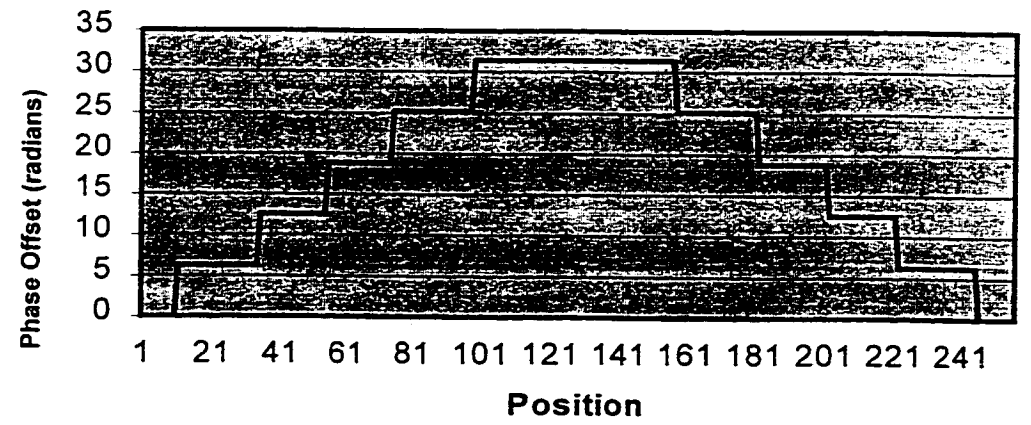


Figure 3.7. One-dimensional Phase Offset Function.

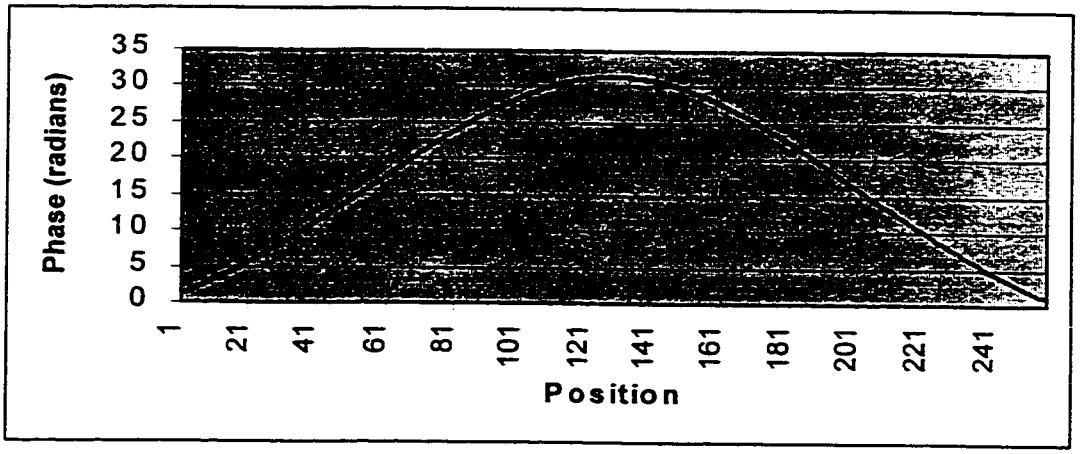


Figure 3.8. One-dimensional Unwrapped Phase.

Automated phase unwrapping represents the crux of the phase recovery problem in holographic interferometry. Existing approaches for the unwrapping of both one- and two-dimensional signals are reviewed in the next two sections.

3.4.1 One-dimensional Phase Unwrapping

The process of phase unwrapping is basically that of numerical integration for which the phase derivative is numerically estimated from the sampled signal and the initial conditions are empirically determined. Unwrapping of a one-dimensional signal involves a search for discontinuities (derivative estimates) greater than a specified threshold. For a continuous phase map, the discontinuity magnitude is 2π radians, but for a sampled signal it can be π radians at the Nyquist sampling frequency. Furthermore, if the image is noisy, as real images are, then the discontinuities may be less than π radians. Takeda *et al.* (1981) suggests a threshold of 0.9π radians. When a discontinuity

exceeding the threshold is found, the sign of the numerical derivative in that neighbourhood is utilised to determine if one cycle (2π radians) must be added to, or subtracted from, the current value of the phase offset. The process is performed until all discontinuities have been removed.

3.4.2 Two-dimensional Phase Unwrapping

The approach adopted by most researchers for unwrapping two-dimensional signals is to simply treat the signal as a sequence of independent, one-dimensional signals. Unwrapping is performed along each row of the sampled phase map, with one final unwrapping along one column to remove discontinuities between the rows (Takeda *et al.*, 1981). The unwrapping can also be performed along each column with the final unwrapping along one row.

Problems with this approach arise when inconsistencies in the phase map are encountered. Inconsistencies may take the form of extra, spurious discontinuities or the absence of discontinuities. The sources of these defects include noise, local aliasing (under-sampling) and discontinuities inherent in the object surface, and hence phase map, topography. Errors due to the inconsistencies are the addition of too many or too few cycles to the wrapped phase map. The resulting phase map possesses a streaked appearance. Moreover, inconsistencies can lead to path-dependent unwrapping. That is, the phase map unwrapped on a row-by-row basis will be different than that produced by integration along a different path, column-by-column for example.

Various attempts to remedy these algorithmic hurdles have been proposed. The approach of Goldstein *et al.* (1988) was to first identify inconsistencies in the phase map based upon residue analysis. A residue is the sum of phase map values over the closed integration path defined by four adjacent samples, the corners of a box (Bone, 1991). If the residue for any such path in the phase map is nonzero, an inconsistency has been identified. One approach to prevent unwrapping errors is the insertion of branch cut lines that an unwrapping integration path can not cross (Goldstein *et al.*, 1988). In a similar approach, Bone (1991) masks inconsistent phase map samples from the unwrapping process and resolves their phase manually. Hellwich (1998) identifies inconsistencies by analysing the difference between the original and a smoothed version of the phase map. In another solution cited by Bone (1991), the phase map is decomposed into smaller regions, each of which are unwrapped. A final unwrapping step is performed to remove the discontinuities between adjacent, unwrapped regions.

The common property of these methods, no matter how inconsistencies are dealt with, is that the two-dimensional unwrapping process is treated as a series of independent, one-dimensional problems. The shortcoming of such an approach is the inability to utilise the strong correlation between neighbouring rows or columns in order to overcome inconsistencies. It will be demonstrated in Chapter Five that gaps in the phase map discontinuities can be bridged and spurious discontinuities eliminated with a new, two-dimensional approach to this problem.

3.5 General Case of Holographic Deformation Measurements

Returning to the task of extracting deformation measurements from a hologram, it will be assumed that a given phase map has been successfully unwrapped. The geometry portrayed in Figure 3.2 is not the most general case, which is treated in this section. Furthermore, the greater resolution resulting from phase map recovery is also treated.

The general case of optical path length difference geometry in holographic interferometry is illustrated in Figure 3.9 (Robinson and Reid, 1993). Here, the deformation is not necessarily in the out-of-plane direction, so angles β and β' (before and after deformation) are not equal. When the object is in its original state, the object wave front propagates along the path described by ABC. The object wave front propagates along the path DEF, though, when the object is in the deformed state. The magnitude of deformation, Δd is given by BE, while the optical path length difference between the two beams, $\Delta \ell$, is described by GEH.

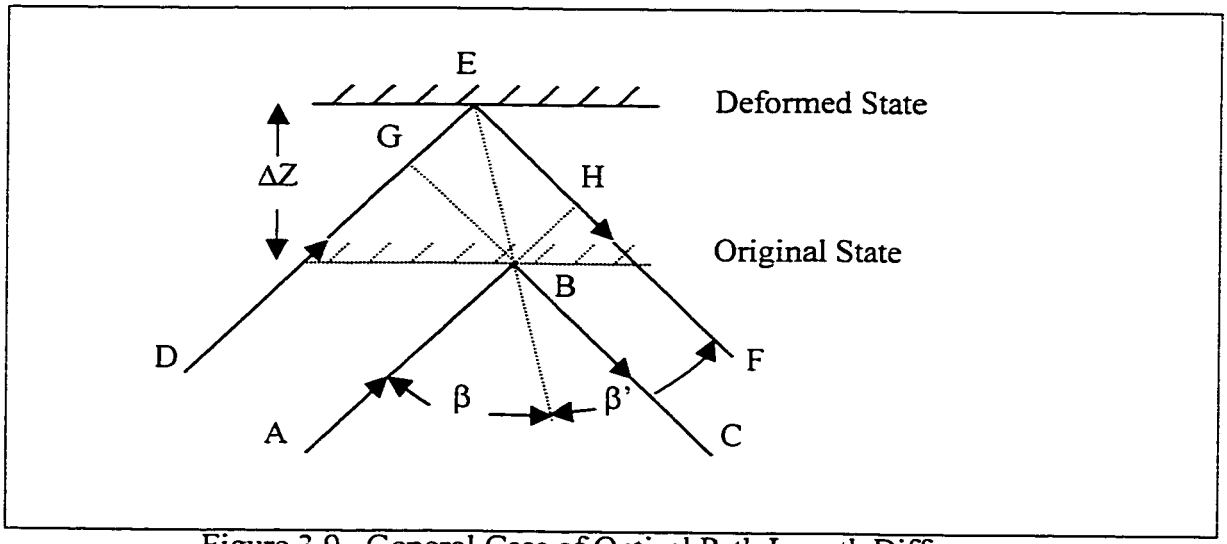


Figure 3.9. General Case of Optical Path Length Difference Geometry in Holographic Interferometry.

From the geometry of Figure 3.9, the path length difference between exposures is equal to

$$\Delta\ell(x,y) = \Delta d(x,y)\{\cos\beta + \cos\beta'\} . \tag{3.27}$$

Substitution of Equation 3.2 into the above expression and rearranging yields the magnitude of the deformation in terms of the phase map

$$\Delta d(x,y) = \frac{\lambda\phi(x,y)}{2\pi\{\cos\beta + \cos\beta'\}} .$$

Using the identity

$$\cos\beta + \cos\beta' = 2\cos\left(\frac{\beta+\beta'}{2}\right)\cos\left(\frac{\beta-\beta'}{2}\right) , \tag{3.28}$$

this expression becomes

$$\Delta d(x,y) = \frac{\lambda\phi(x,y)}{4\pi\cos\left(\frac{\beta+\beta'}{2}\right)\cos\left(\frac{\beta-\beta'}{2}\right)} . \tag{3.29}$$

For the case of pure, out-of-plane deformation, $\beta = \beta'$, and the deformation magnitude, $\Delta d = \Delta Z$

$$\Delta Z(x,y) = \frac{\lambda\phi(x,y)}{4\pi\cos\beta} . \tag{3.30}$$

In the case of on-axis holography, the phase map consists of integer multiples of 2π, and Equation 3.30 reduces to

$$\Delta Z(x,y) = \frac{\lambda N_i}{2\cos\beta} . \tag{3.31}$$

Comparison of the above expression with Equation 3.4 indicates that the incidence angle between the surface normal and object beam, β, reduces the resolution of

deformation measurements by a factor of $\cos\beta$. If, for example, $\beta = 30°$, then the resolution is reduced from 0.5λ to approximately 0.577λ, a loss of approximately fifteen per cent. In off-axis holography, the lowest resolution of phase map recovery is π radians, which occurs at the Nyquist frequency. This translates into a deformation measurement resolution of 0.25λ. Thus, by directly recovering the phase map, not only is the concavity ambiguity removed, but also the resolution of out-of-plane deformation measurement is improved by a factor of two.

3.6 Chapter Summary

The purpose of this chapter was to demonstrate different methods for the extraction of a phase map and, therefore, a deformation field, from different types of holographic imagery. This review highlighted some of the pertinent issues with the various phase recovery methods, such as the ambiguity and unwrapping problems. This represents a logical basis for the development of new algorithms to solve these problems in Chapters Four and Five.

CHAPTER 4

FRINGE MAXIMA EXTRACTION

Analysis of a HI intensity image, as opposed to the underlying phase map, is often used for industrial inspection and vibration studies. For these applications, only the absolute value of peak displacement and/or mode shapes is required. The task in such analyses is to extract the peaks or ridges in sinusoidal intensity from the interferogram. Due to speckle noise and non-linear fringe visibility variation over an image, simple local maxima detection has proven to be insufficient (Robinson and Reid, 1993). A new approach, which offers a significant improvement in overcoming these problems, exploits the topology of fringe imagery and utilises various image-processing methods.

4.1 Theoretical Basis

As detailed in Chapter Three, each successive fringe maxima of a holographic interferogram represents a change in phase of 2π radians. As expressed as contour lines, fringe maxima must obey certain rules. First, they must be closed, simple curves that can not cross over themselves or over other contours. An open contour line is also valid provided that it starts and ends at the border of a finite image. Special cases of contour lines are also valid. These include ridgelines and peaks (points). Such information may

be crucial in determining peak displacement due to vibration. A region is defined as an area of variable intensity bounded by fringe maxima contours. Regions are not necessarily simply connected, as they may contain holes, which are other regions.

A dual network of contour lines also exists in fringe imagery. Fringe minima (dark fringes) contain dual information about the topography of the underlying phase map. The fringe minima only differ by π radians (one-half cycle) from the maxima.

An important consideration in the analysis of discrete fringe imagery is the sampling interval. The minimum interval is two samples per fringe in each dimension. However, an oversampled image is more desirable so that the reproduced fringes are clearly defined. High sampling rates may also be required by the fact that the spatial frequencies of the fringes are not constant. In the following developments, it will be presumed that the interferogram is adequately sampled.

The topological description of a discrete, interference image begins with the definition of the interferogram, I. An interferogram is a two-dimensional, finite region comprised of m regions, R_i, $i = 1..m$, and n maxima contours, C_j, $j = 1..n$. Using set theory, the topological rules of the inteferogram are expressed as

$$R_i \cap R_j = \varnothing \qquad \forall i, j = 1..m, i \neq j \ , \tag{4.1a}$$

$$C_i \cap C_j = \varnothing \qquad \forall i, j = 1..n \ , \tag{4.1b}$$

$$R_i \cap C_j = \varnothing \qquad \forall i = 1..m, j = 1..n \ , \tag{4.1c}$$

$$\left(\bigcup_{i=i}^{m} R_i \right) \cup \left(\bigcup_{j=i}^{n} C_j \right) = I \ . \tag{4.1d}$$

Equation 4.1a states that the intersection of two regions equals the null or empty set. That is, no two regions of the interferogram may overlap. Similarly, Equation 4.1b states that no two contours, which are not of infinitesimal but discrete width, may cross and that no contour can cross over itself. Equation 4.1c states that no contour can occupy any sample of the interferogram belonging to a region, and vice-versa. Finally, Equation 4.1d states that the interferogram, I, is comprised of the union of all contour and region samples. Thus, gaps in the image are not permitted. As a consequence of defining these rules, any given contour can only bound two regions (one on either side). However, a region may be bounded by more than one contour. With the topological rules formally defined an algorithm for contour-based fringe maxima extraction from noisy imagery can be described.

4.2 Description of the New Method

The proposed method for fringe maxima extraction consists of several steps, which are summarised in the following list.

1. Homomorphic filtering.
2. Thresholding.
3. Thinning.
4. Contour location and following.
5. Contour linking and spur removal.
6. Artefact identification and rectification.

The purpose of the homomorphic filtering is to reduce the multiplicative speckle noise. By thresholding the filtered image, the areas of peak intensity (white fringe maxima) are retained while the minima are discarded. Thinning is then performed to reduce the broad fringe maxima to single pixel-wide contour lines. The thinned contours are then located and tracked (followed) through the image. Using the topological rules of fringe imagery, spurious contours are removed and gaps bridged. Finally, additional measures are required to overcome severe distortions in the contours so that the topological rules are satisfied. Such artefacts arise from the speckle noise and the filtering.

4.2.1 Homomorphic Filtering

The role of homomorphic filtering, as detailed in Chapter Two, is to reduce the multiplicative speckle noise in a HI image. In tests conducted with real imagery, homomorphic filtering was found to be superior, as determined by visual inspection, to the simpler low-pass filtering option for speckle reduction. Since the spectral content of on-axis holographic imagery is clustered at low spatial frequencies, Butterworth low-pass filters were chosen for the homomorphic filtering. This choice was motivated by the Butterworth's low attenuation at low frequencies and narrow transition band. The spatial-frequency domain representation of a circularly symmetric, Butterworth low pass filter of order n is given by

$$H(\omega,\nu)=\frac{1}{1+\left(\frac{\sqrt{\omega^2+\nu^2}}{r_c}\right)^{2n}}, \tag{4.2}$$

where

r_c is the radial cut-off frequency.

As will be shown in Chapter Six, there is significant improvement in HI image quality as a result of the homomorphic filtering. The reduction of speckle produces smoother contours and fewer spurious contours in later stages of the algorithm. However, the blurring caused by the attenuation of high frequencies creates false contours. Although this distortion creates the need for step six of the algorithm, artefact identification and rectification, the improvement in image quality outweighs this drawback.

4.2.2 Thresholding

In order to extract the fringe maxima from a filtered HI image, the image is segmented via thresholding. Intensity values above the threshold, the fringe maxima, are assigned a value of one, while those values below the threshold are assigned a value of zero. Two practical problems complicate the use of such a simple segmentation rule. First, the choice of the threshold for HI imagery is often not obvious. Rather than possessing a bimodal histogram with clearly defined peaks, the imagery used in this study had either a poorly defined second peak, or possessed only a single peak. Furthermore, the homomorphic filtering appeared to transform the histograms into unimodal shapes. Examples taken from real HI imagery will be presented in Chapter Six to illustrate these problems. The solution adopted was to set the threshold at the centre of mass of the histogram, as defined by

$$T = \frac{\sum_{n=0}^{Q-1} n \cdot f(n)}{\sum_{n=0}^{Q-1} n}, \tag{4.3}$$

where

n is the intensity value,

f(n) is the frequency of occurrence of intensity value n,

Q is the number of possible intensity levels (i.e., 256).

The second difficulty that arises in thresholding HI imagery stems from non-linear fringe visibility. As will be illustrated in Chapter Six, such variation makes the application of a global threshold inappropriate. Instead, the imagery is decomposed into smaller sub-images. Thresholding is then performed locally within each sub-image.

4.2.3 Thinning

Thresholding of the filtered imagery, does not, unfortunately, yield single pixel-wide contour lines. Instead, broad fringes are the products of this process. Binary image morphology, which is the analysis of shapes in images (Gonzalez and Woods, 1992), provides a means of thinning these fringes through a sequence of non-linear filtering operations.

The two basic morphologic operations are dilation and erosion. The dilation operation causes features within an image to grow. The dilation of a feature or set, A, with a structuring element, B, is denoted by

$$A \oplus B \ . \tag{4.4}$$

Like in discrete convolution, the structuring element (mask) is reflected and translated across the entire image. Unlike convolution, though, the filter response at any location is not the weighted sum of intensity values in a neighbourhood. Instead, the response is calculated with logical operations (Castleman, 1996). For dilation, the response is equal to one if at least one sample of A is nonzero within the neighbourhood defined by the structuring element. Otherwise, the response is equal to zero.

Erosion can be thought of as the dual operation to dilation (Gonzalez and Woods, 1992). Denoted by

$$A \ominus B \ . \tag{4.5}$$

the erosion operation has the effect of eroding or shrinking features in an image. The erosion response is equal to one if all the samples of A that lie within the neighbourhood of B are nonzero. Otherwise, the response is equal to zero.

The hit-or-miss transform (HOMT) is a non-linear combination of these basic operations that is an integral part of the thinning operation. The HOMT is a shape detection operation used to locate features such as corners (Haralick and Shapiro, 1992). The HOMT of a set, A, uses a structuring element, J, and its complement, K, where

$$K = W - J \ . \tag{4.6}$$

where W is the local background of the structuring element. The HOMT operation is defined as

$$A \otimes J = (A \ominus J) \cap (A^c \ominus K) \ . \tag{4.7}$$

Analysis of this equation indicates that the HOMT produces the intersection of eroded versions of A and its complement, A^c.

The thinning operation utilises the HOMT and is implemented as

$$A \otimes J = A - (A \circledast J) \ . \tag{4.8}$$

Thus, the thinned version of A is simply the set A, itself, minus the output of the HOMT. Thinning is generally implemented recursively using a set of structuring elements. That is, the output of one thinning operation becomes the input set for the next.

4.2.4 Contour Location and Following

Once the fringe maxima have been reduced to single pixel-wide contours, they can be extracted from the imagery. First, the image is searched for a nonzero sample. Once one is found, its position is recorded and its neighbours are analysed. If a second nonzero sample is found, its position is recorded and its neighbours examined. Only the adjacent neighbours within ±45° of the direction from the previous to current samples are examined. This process is repeated until either the image border or the end of the contour is reached. The result of this process is a chain code for each contour describing its spatial position in the imagery.

4.2.5 Contour Linking and Spur Removal

Having obtained a spatial description of each contour in a HI image, the topological properties of fringe imagery can be enforced in order to edit the contours. Knowing that gaps are not permitted in contours, gaps between neighbouring contour ends can be bridged with short line segments. The gap bridging is restricted by the condition that the difference in azimuth between the two tangents can not exceed a pre-

defined tolerance. This constraint ensures that two parallel contours are not erroneously joined.

Another artefact arising from the distortions inherent in HI imagery is a spur contour. A spur is a short contour that abuts two others at a T-intersection. An example of a spur is the curve BD in Figure 4.1a.

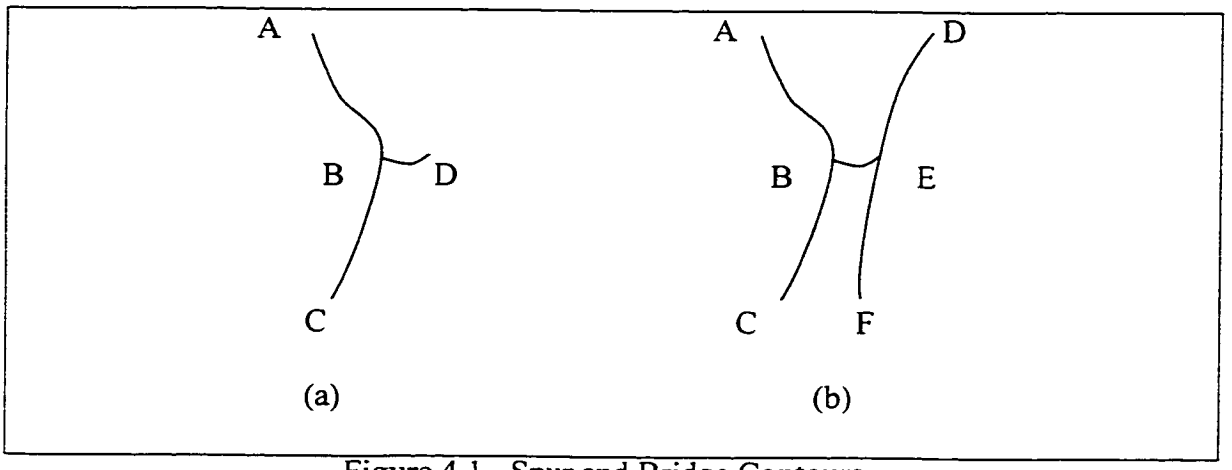


Figure 4.1. Spur and Bridge Contours.

Using contour length as the decision criterion, BD is eliminated if its length is less than that of both of contours AB and BC. This results in a single contour, AC.

A similar artefact is a bridge contour. A bridge, as denoted by curve BE in Figure 4.1b, is basically a spur contour at both ends. If the length of BE is less than all four of AB, BC, DE and EF, then BE is eliminated. After elimination of the bridge, only two nominally parallel contours, AC and DF, remain.

4.2.6 Artefact Identification and Rectification

Even after the automated gap, spur and bridge removal procedures have been performed, some erroneous contours and gaps will exist. The number of these residual artefacts chiefly depends upon the severity of the speckle noise and filtering distortions. In order to rectify these problems, three semiautomatic procedures are proposed below.

4.2.6.1 Linear Contour Interpolation

The first type of artefact considered is what might be termed an X-junction of contours. Figure 4.2a illustrates this artefact in which four contours, AE, BE, CE and DE, converge at the central point, E.

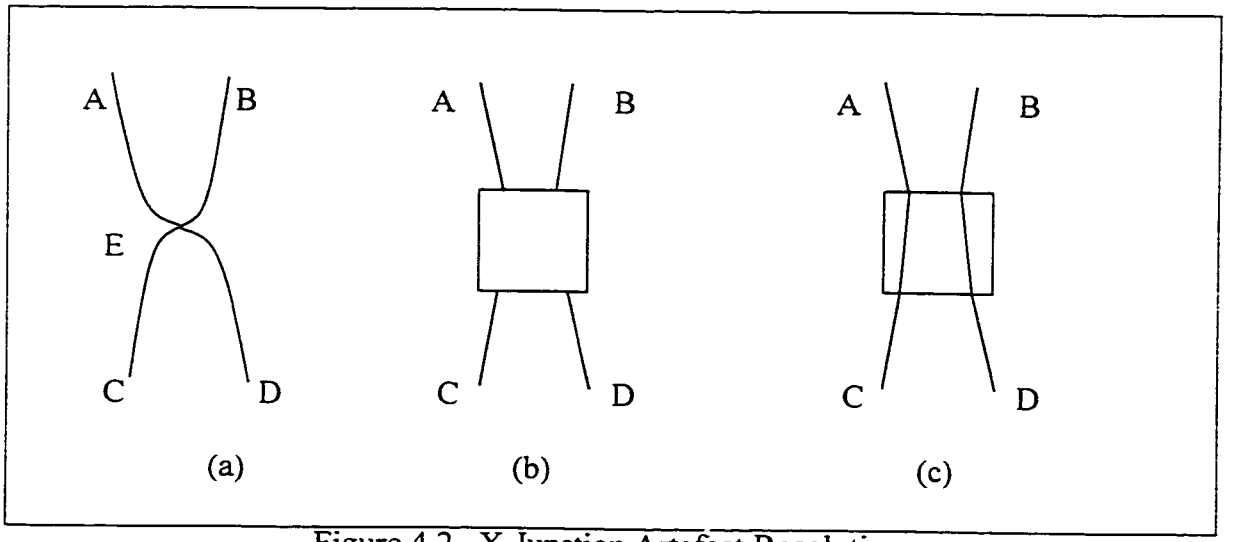


Figure 4.2. X-Junction Artefact Resolution.

Upon visual inspection, such a problem appears trivial to resolve based upon the trend exhibited by each contour line. Nevertheless, this artefact is difficult to automatically correct. A semiautomatic solution begins with interactive identification of the artefact.

This step, shown in Figure 4.2b, simply consists of placing a window around the neighbourhood of the X-junction. The portion of each contour within the window is then eliminated. The problem is then automatically resolved by using the procedure for gap filling described in Section 4.2.4. The linearly interpolated contours are shown in Figure 4.2c.

4.2.6.2 Concentric Contour Generation

Another artefact that is easily resolved in semiautomatic mode is an incomplete closed contour. Illustrated in Figure 4.3a is a complete, closed contour encircled by an incomplete contour, AB. The goal here is to bridge the gap between A and B, which may contain distorted contour fragments.

As with the linear interpolation procedure, the first step in concentric contour matching is identification of the artefact (a gap, in this case) area. The constructed window is shown in Figure 4.3b. Obviously, straight-line interpolation is inappropriate due to the curvature of the contour. Instead, the inner closed contour is exploited. Using points on the extracted portion of contour AB, a scale factor is estimated which maps the closed contour onto AB. A portion of the scaled version of the closed contour, illustrated by the dashed curve, CD, in Figure 4.3c, is then constructed within the window area. Gaps in the scaled contour segment are filled via linear interpolation. In order to match the endpoints C to A and D to B, a similarity transformation is performed on curve CD. The final closed curve is shown in Figure 4.2d.

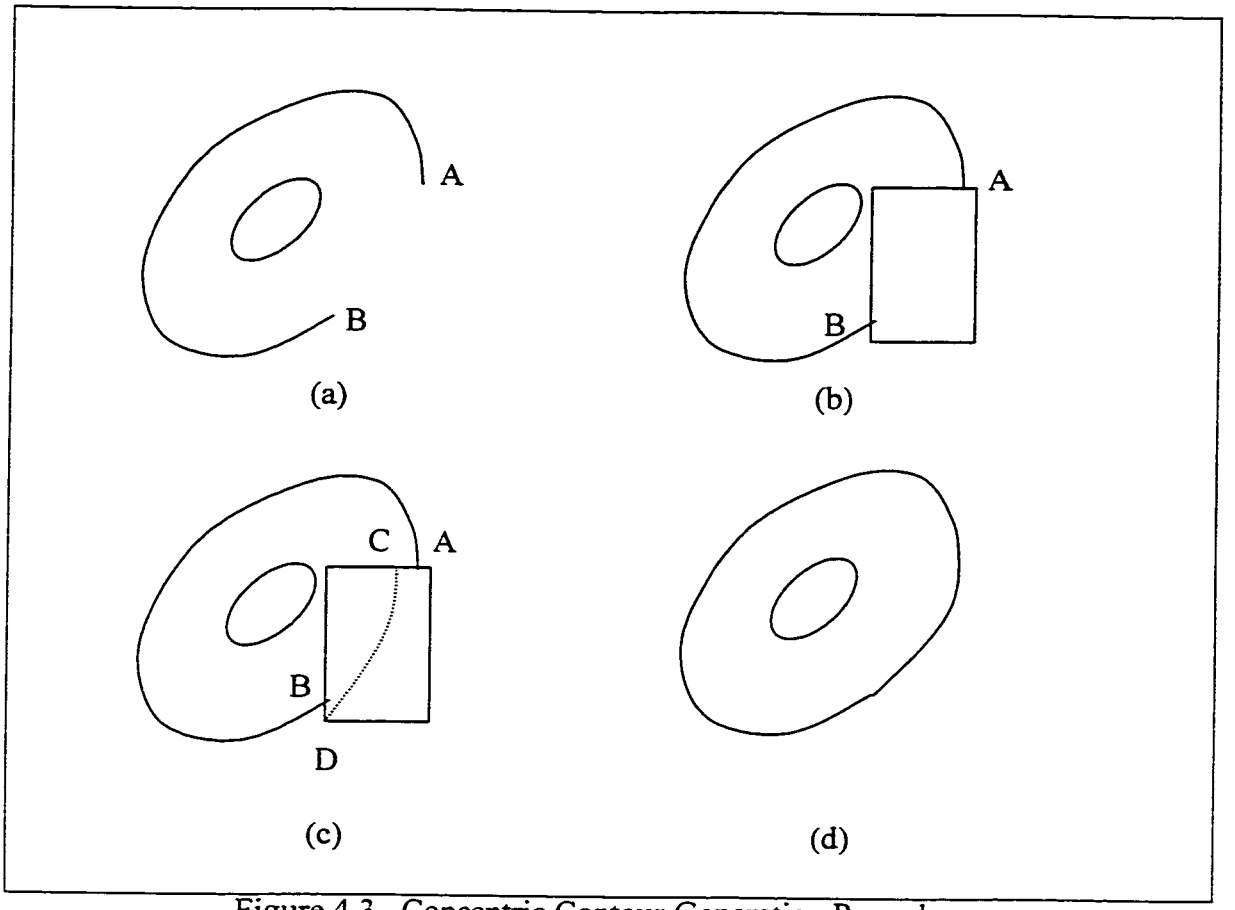


Figure 4.3. Concentric Contour Generation Procedure.

The success of this method hinges on two factors: the smoothness (or roughness) of the inner contour, and the contour shape consistency. Although any gap can be bridged with this matching technique, the roughness of the inner contour may yield an inappropriate result. Due to the scaling operation, the roughness of the closed contour is amplified, which may cause the interpolated segment to significantly deviate from its true position. This situation is illustrated in Figure 4.4. The rough, inner contour and incomplete contour are shown in Figure 4.4a, while the result of the interpolation is shown in Figure 4.4b.

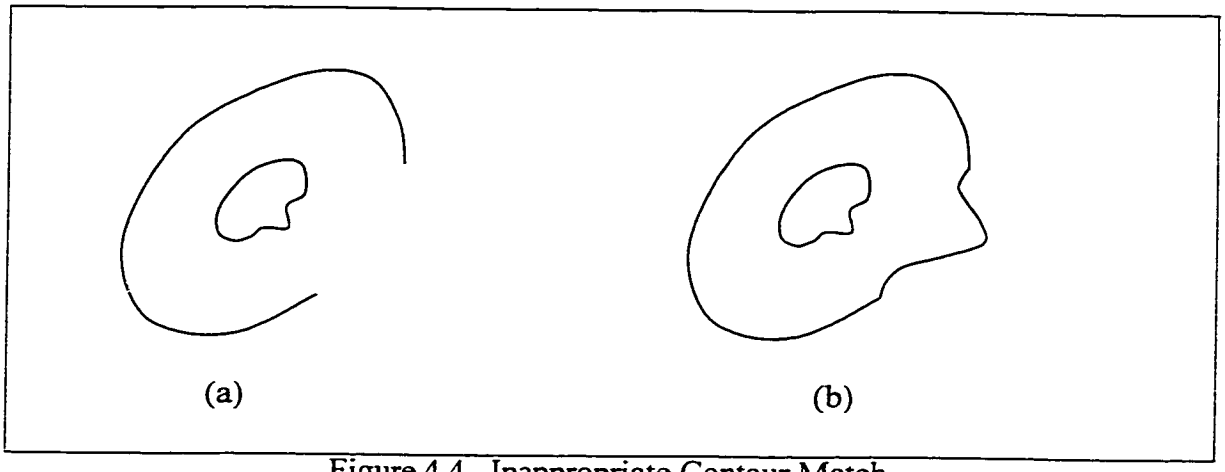


Figure 4.4. Inappropriate Contour Match.

A similar problem is contour shape inconsistency. This occurs when the shape of the inner contour is not representative of the true shape of the incomplete contour to be interpolated. Judgement must therefore exercised by the user when applying this technique.

4.2.6.3 Quadratic Contour Interpolation

The final artefact resolution procedure to be described requires slightly more interaction. For situations in which linear interpolation and concentric contour matching are inappropriate, the user must supply additional information. In order to bridge the gap between A and B in Figure 4.5a, the endpoints, A and B, are first identified by the user. The user also provides an intermediate point, C, through which the interpolating curve will pass. These three points, shown in Figure 4.5b, uniquely define a parabolic curve parameterised as

$$y = a_0 + a_1 x + a_2 x^2 \tag{4.9a}$$

or

$$x = a_0 + a_1 y + a_2 y^2 \ . \tag{4.9b}$$

The choice of parameterisation is dependent upon the separation between the endpoints in each co-ordinate direction. Having estimated its coefficients, the parabola is then used to bridge the gap between A and B. The completed contour is shown in Figure 4.5c.

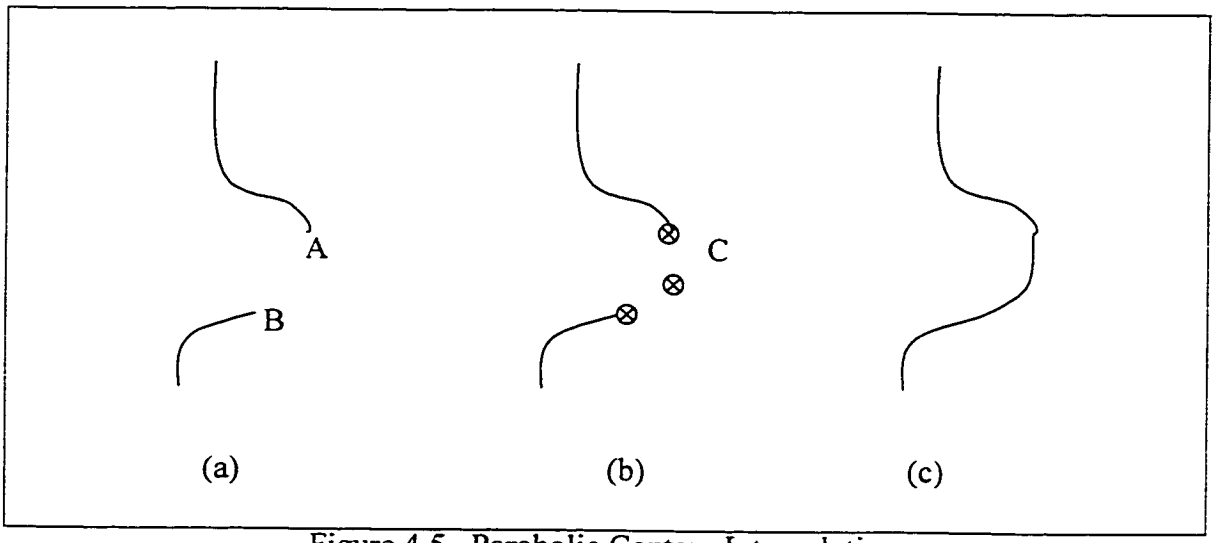


Figure 4.5. Parabolic Contour Interpolation

In areas where contour behaviour is not adequately described by a single parabola, two options are available. The first possibility is to employ a higher-order method of interpolation, such as a cubic polynomial. The second possibility is to perform piecewise interpolation with quadratic polynomials. The latter option was exercised in this research.

4.3 Chapter Summary

A new approach to fringe maxima contour extraction has been outlined in this chapter. The foundation of the method is a set of topological rules governing the relationships between various entities in a fringe image. The algorithm is comprised of a series of image processing, computer vision and editing steps that are portrayed in the flowchart in Figure 4.6.

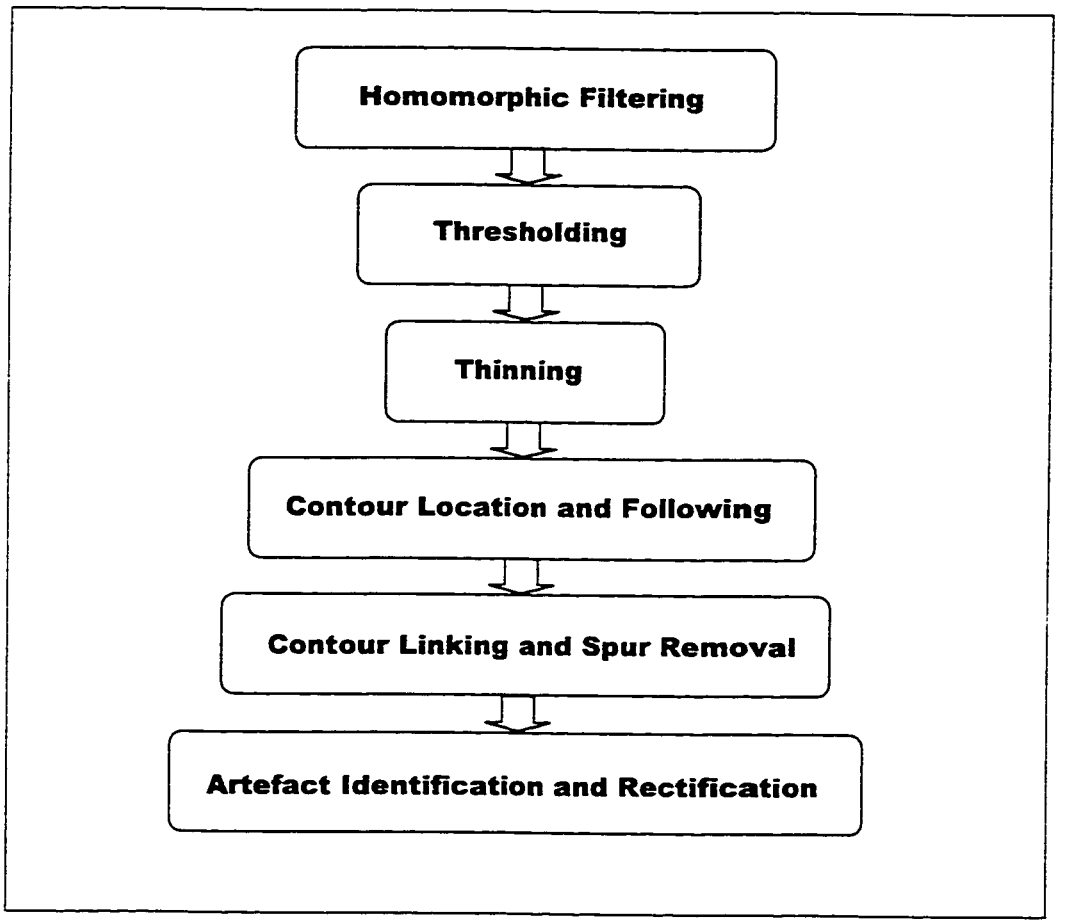


Figure 4.6. Fringe Maxima Contour Extraction Algorithm.

CHAPTER 5

CONTOUR-BASED TWO-DIMENSIOANL PHASE UNWRAPPING

As mentioned in Chapter Three, the major drawback of current two-dimensional phase unwrapping algorithms lies in their nature as a series of independent, one-dimensional unwrapping problems. Because each row (or column) is unwrapped independently of its neighbours, any spurious discontinuities or gaps in the discontinuities will cause unwrapping errors Goldstein *et al.* (1988).

A phase map represents a surface that is proportional to the topography of the measured object. Because of this property, it intuitively makes sense to treat a phase map as a topographic surface as such rather than a collection of independent, one-dimensional signals. A new, contour-based method that overcomes the aforementioned unwrapping errors by utilising the information contained in adjacent rows and/or columns is presented in this chapter.

5.1 Theoretical Basis of the Contour-Based Method

In the new two-dimensional phase unwrapping algorithm, a wrapped phase map is defined to be a discontinuous surface comprised of contours and regions. A contour is a simple, closed curve described by the locus of points of a 2π discontinuity. Contours are

isolines and, therefore, must not cross over themselves or over other contours. In fact, contours, as defined here, are nothing more than the 2π contour lines of the unwrapped phase map. Since wrapped phase maps have finite support, not all contours will necessarily be closed. An open contour is a curve that both starts and ends at the image boundaries. A hanging contour, however, is a curve for which the start and/or end are located at a position within the image interior. These curves are caused by noise and distortions and are not valid phase-map contours.

A region is defined as an area of continuous phase bounded by contours. The amplitude of a region is bounded on $[-\pi, \pi]$ radians due to the wrapping. Regions may not necessarily be simply connected, as the may contain holes (other regions). Regions and the different types of contours are illustrated in Figure 5.1.

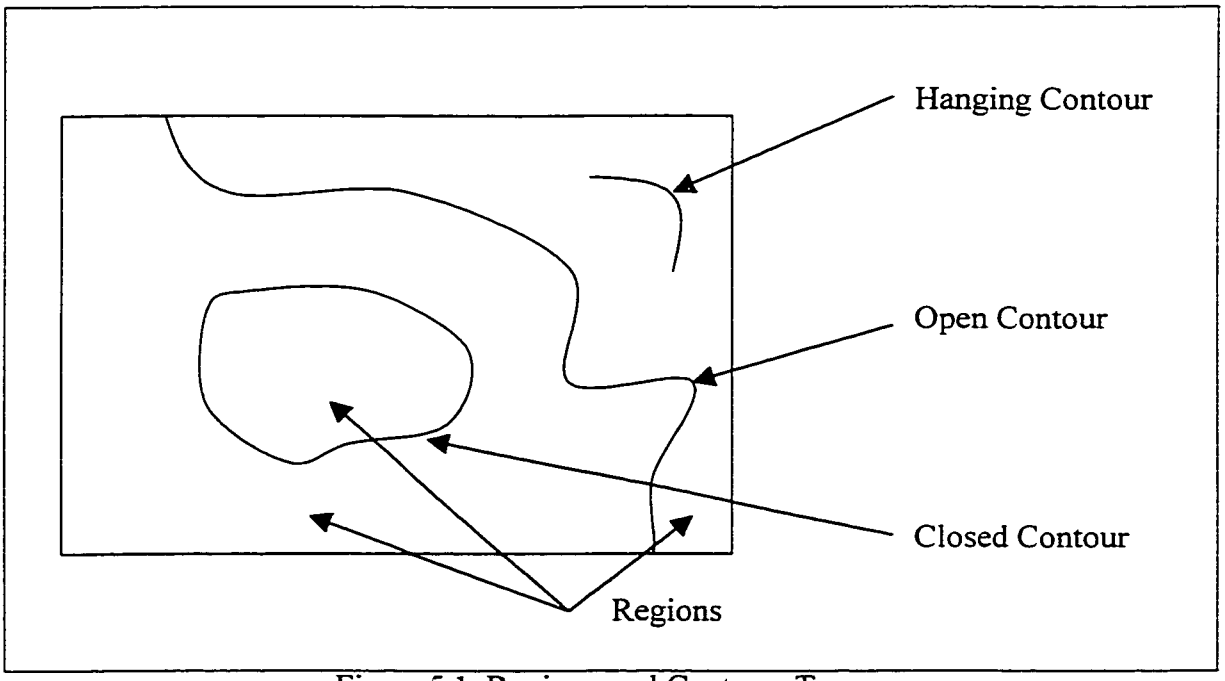


Figure 5.1. Regions and Contours Types.

The contours and regions of the wrapped phase map in Figure 4.4 are given in Figure 5.2. Each of the ten regions is delineated by a different shade of grey. The nine contours are marked in white.

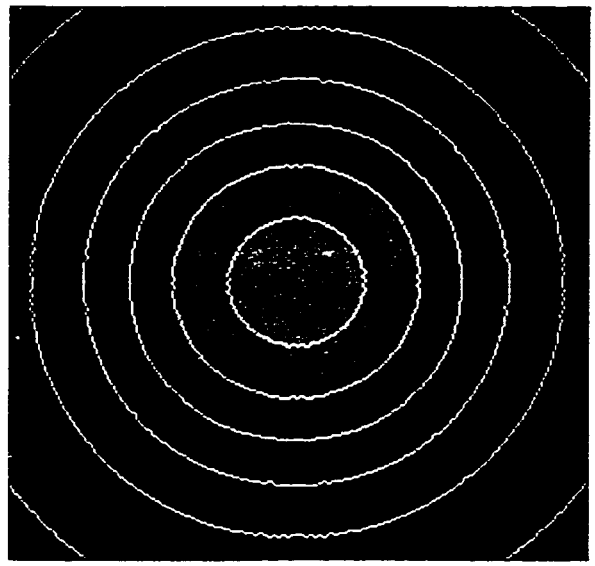


Figure 5.2. Regions and Contours of the Wrapped Phase Map.

In order to define the relationships between phase map entities (contours and regions) for the purpose of unwrapping, it is first necessary to impose two conditions on the sampled phase map. First, the wrapped phase map must satisfy the Nyquist sampling criterion that at least two samples per region and the relative phase offset between two adjacent regions is 2π radians. This assumption must be taken in the context of global aliasing. In the next chapter, it will be shown through simulation that local aliasing of a region can be overcome with this method. Second, the original, unwrapped phase must be a continuous surface. That is, the surface can not be fractured or stepped.

These assumptions may appear somewhat restrictive, but are in fact required for most unwrapping algorithms, whether stated explicitly or not. There are notable exceptions, though. Greivenkamp (1987) details a method in which derivative constraints are used to determine the most likely number of cycles to add to aliased areas of a phase map. However, the form of the resulting phase map is dependent upon the placement of an initial seed location from which unwrapping commences and requires continuity of the surface and its derivatives. This method is also one-dimensional as it is performed on a row-wise or column-wise basis. Bone's (1991) algorithm circumvents the surface continuity requirement by *a priori* identification and masking of discontinuities in the phase map.

The topological description of a discrete, wrapped phase map begins with the definition of the phase map, P. A phase map is a two-dimensional, finite region comprised of m regions, R_i, $i = 1..m$, and n contours, C_j, $j = 1..n$. Using set theory, the topological rules of the phase map are expressed as

$$R_i \cap R_j = \varnothing \qquad \forall i, j = 1..m, i \neq j \ . \tag{4.1a}$$

$$C_i \cap C_j = \varnothing \qquad \forall i, j = 1..n \ , \tag{4.1b}$$

$$R_i \cap C_j = \varnothing \qquad \forall i = 1..m, j = 1..n \ , \tag{4.1c}$$

$$\left(\bigcup_{i=i}^{m} R_i \right) \cup \left(\bigcup_{j=i}^{n} C_j \right) = P \ . \tag{5.1}$$

Equation 4.1a states that the intersection of two regions equals the null or empty set. That is, no two regions of the phase map may overlap. Similarly, Equation 4.1b states that no two contours, which are not of infinitesimal but discrete width, may cross and that

no contour can cross over itself. Equation 4.1c states that no contour can occupy any sample of the phase map belonging to a region, and vice-versa. Finally, Equation 5.1 states that the phase map, P, is comprised of the union of all contour and region samples. Thus, gaps in the phase map are not permitted. As a consequence of defining these rules, any given contour can only bound two regions (one on either side). However, a region may be bounded by more than one contour. Note the similarity between these rules and those of fringe images presented in Section 4.1. With the topological rules formally defined, the algorithm for contour-based phase unwrapping can be described.

5.2 Description of the Contour-Based Method

The proposed contour-based phase unwrapping algorithm consists of several steps that are summarised in the following list.

1. Edge detection.
2. Edge location and following.
3. Edge linking.
4. Region identification.
5. Contour-wise unwrapping.

The role of edge detection is to simply locate significant discontinuities in a phase map through convolution with a bank of suitable filters. The aim of edge location and linking is to obtain a spatial description, in the form of a chain code, of each contour. In the edge linking step, gaps between hanging contours and spurious, hanging contours are eliminated. The purpose of region identification is to obtain a spatial description of each

region. Finally, in the contour-wise unwrapping step, the relationship between regions is established using the contour gradient information.

5.2.1 Edge Detection

An edge in a digital image is a discontinuity in intensity, which may be blurred due to filtering operations and is invariably contaminated by noise. An edge in a wrapped phase map also follows this definition. The task of edge detection is location of these discontinuities.

Since discontinuities represent large intensity gradients, numerical differentiation can be used as an edge detection tool. However, the differentiation operation enhances high frequency noise, as evident in the linear modulation of the spectrum in Equation A.5 (Hamming, 1989). This can greatly inhibit the success of edge detection since differentiation of a noisy signal yields an even noisier one. If, however, the high frequency components of the image are attenuated beforehand via low pass filtering, the noise amplification effect is greatly reduced. This benefit comes at the expense of edge blurring. The smoothing and detection operations can be combined into one. Because convolution with a filter and differentiation are both linear operations, their sequential application is equivalent to convolving with the first derivative of the filter impulse response, if it exists.

A low pass filter approximation often used for image processing is the circularly symmetric Gaussian function. Thus, a smoothing differentiation filter is the first derivative of the Gaussian, described in Section A.7. This filter was proposed by Canny

(1986) as a close approximation to an optimal step edge detector. However, the general form of phase map edges is not a step, but more closely resembles a triangular or saw tooth edge. Because of its smoothing properties, though, this filter is preferred to edge detection kernels such as the Sobel and Prewitt masks.

As detailed in Appendix A, a multi-scale image representation can be used to capture prominent details at different scales. That is, high frequency details can be captured at small scales while low frequency trends reveal themselves well at large scales. For wrapped phase map imagery, different operator widths can be used to capture edges bounding regions of different widths. The primary advantage of this approach, though, is the noise reduction achieved at larger filter widths. However, spatial aliasing will result if the filter width is too great due to the contributions of neighbouring edges to the filter response.

The first stage of edge detection for phase map unwrapping is convolution with the kernels, given by Equations A.22 and A.23, at different scales. This yields a family of scale-dependent images for both x- and y-gradients:

$$g^x(x,y,\sigma) = \psi^x(x,y,\sigma) ** \phi_d(x,y) \tag{5.2a}$$

$$g^y(x,y,\sigma) = \psi^y(x,y,\sigma) ** \phi_d(x,y) \ , \tag{5.2b}$$

where

$g^x(x,y,\sigma)$, $g^y(x,y,\sigma)$ are the scale-dependent gradient images in the x- and y-axis directions, respectively,

$\psi^x(x,y,\sigma)$, $\psi^y(x,y,\sigma)$ are the multi-scale kernels given by Equations A.22 and A.23, respectively.

Since noise characteristics may vary, the number of different scales required for a given phase map may also vary. The parallel structure of multi-scale image representation is adopted here, thereby retaining the original image dimensions at each scale (Schenk, 1995).

Having calculated the responses at each scale, the next step is their fusion. The purpose of the fusion is to combine the responses such that the edge information contained within each is fully exploited. In the squared gradient approach (Shao and Förstner, 1994), the sum of squares of the magnitude response at each scale is calculated as

$$g(x,y)=\sum_{i=1}^{S} w_i \left\{ g^x\left(x,y,2^i\right)^2 + g^y\left(x,y,2^i\right)^2 \right\} \quad . \tag{5.3}$$

where

S is the number of scales to be fused,

w_i is the weight of the magnitude response at each scale,

$2^i = \sigma$ as defined by Equation A.21.

The key to the success of fusion lies in the relative weighting of the various responses. Fraser and Shao (1996) derive a weight that is inversely proportional to the range of response at each scale.

For this application, a different scheme is required since the response components (gradient orientation) must be preserved for the unwrapping step. The scheme employed here is to separately add the weighted x- and y-responses from each scale:

$$g^x(x,y)=\sum_{i=1}^{S} w_i g^x\left(x,y,2^i\right) \tag{5.4a}$$

$$g^y(x,y)=\sum_{i=1}^{S} w_i g^y\left(x,y,2^i\right). \tag{5.4b}$$

Unity weighting factors were applied. From the fused responses, the edge magnitude, $g(x,y)$, and orientation, $\theta(x,y)$, can be calculated as (Lam, 1990)

$$g(x,y)=\sqrt{g^x(x,y)^2+g^y(x,y)^2} \tag{5.5a}$$

$$\theta(x,y)=\arctan\left(\frac{g^y(x,y)}{g^x(x,y)}\right). \tag{5.5b}$$

5.2.2 Edge Location and Following

Edge following is a computer vision technique used to track edges and lines in imagery for the purpose of segmentation into distinct regions. A precursor to this step, edge location, entails searching the fused response image for local maximum responses. In order to simplify the edge following, the imagery was searched in an inward spiral pattern. The edge location process produces start points (seeds) for the edge following.

From a seed point, a search of three neighbouring samples is conducted for candidate edge pixels or nodes. The chosen pixel is the one with the largest edge magnitude surpassing a threshold. In addition, the orientation of the candidate pixel must

not differ from that of the current edge pixel by ±90°. This criterion ensures that edge tracking does not cross over saddle points where phase map contours pass close by each other. Once the next node has been selected, a search is then performed in its three neighbouring pixels. The neighbours chosen are those within ±45° of the direction from the previous node to the current node, as shown in Figure 5.3. The edge following process terminates when the start point is reached, an image boundary is reached, or a suitable candidate can not be found.

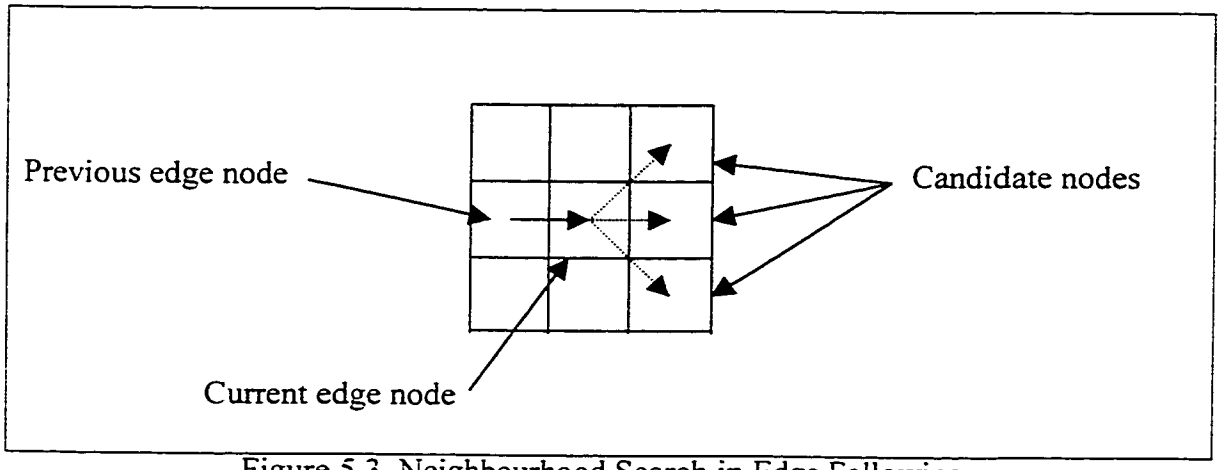


Figure 5.3. Neighbourhood Search in Edge Following.

The position, edge magnitude and edge orientation of each node is recorded in a list structure. The result of edge following is a set of chain codes describing the spatial position of the phase map contours. The contours are then classified according to their start and end points as closed, open or hanging.

5.2.3 Edge Linking

The purpose of fusing filter responses is to bridge gaps in detected contours at small scales with those detected at larger scales. The noise that creates gaps at low scales is greatly reduced by the smoothing effect of broader filters. As a result, the fused edges, although broader, are more continuous in magnitude.

However, some fused edges will fall below the edge following threshold, resulting in contour line gaps. Edge linking is an attempt to merge hanging contours separated by small gaps. The approach of Castleman (1996) is an outward search in the five by five neighbourhood of an endpoint. The algorithm used here is similar. The list of contours is searched for hanging contours. For the first such contour encountered, the remainder of the list is traversed, again searching for hanging contours. For each one found in the second search, the co-ordinate differences for all four end point combinations are calculated. If one set of differences is less than a two-pixel threshold, the two contours are merged into one. This merged contour is then classified. This process continues until no hanging contours remain, or no further merging is possible within the specified tolerance. Finally, any remaining hanging contours are deemed to be spurious and are discarded. The result of this process is a set of chain codes of only closed and open contours.

5.2.4 Region Identification

Having compiled a description of each contour of the phase map, the next step is to obtain a description of the regions. By performing region growing outward from the

contour nodes, the pixels in each region are assigned an identification number. This process establishes the topology of the phase map in terms of which region lies on either side (left or right) of each contour. However, the height topology (phase offset) of each region is not yet established. This is done in the unwrapping step.

5.2.5 Contour-wise Unwrapping

The final step of the contour-based phase unwrapping algorithm consists of two stages, unwrapping of the regions and unwrapping of the contours. The region unwrapping is further broken down into relative and absolute unwrapping. The contour unwrapping is necessary because of the discrete nature of the sampled phase map – contours with finite width.

In relative unwrapping, the relative phase-offset is established for the pair of regions lying on either side of each contour. The goal of this task is to ascertain which region is higher, the left or the right. Edge direction is utilised for this purpose. As shown in Figure 5.4, the relative topology of two regions depends upon the edge direction, denoted by the vectors. In Figure, 5.4a, the direction of the vectors indicates that the left region, L, is one cycle higher than the right region, R. Conversely, the vector directions in Figure 5.4b indicate that the right region is higher than the left.

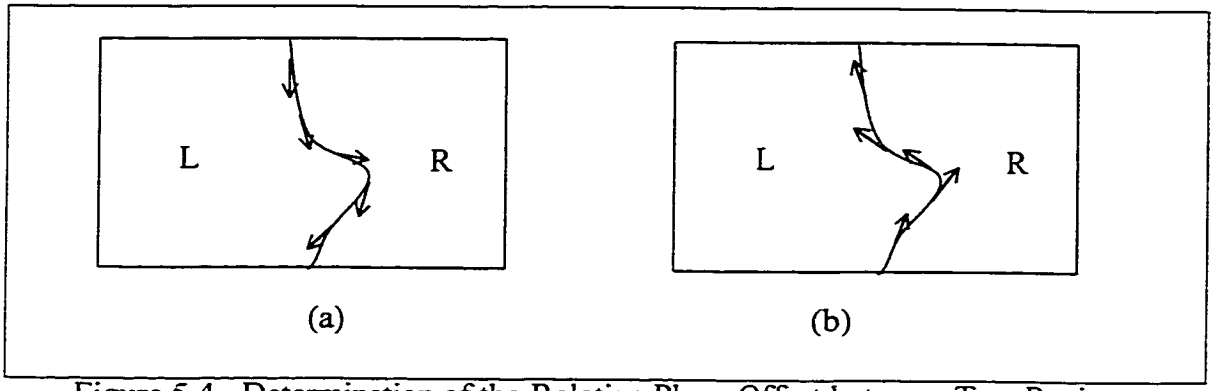


Figure 5.4. Determination of the Relative Phase Offset between Two Regions.

For each contour node, the edge direction is analysed and the decision is made as to which region is higher. The region with the greater frequency of occurrence (number of nodes) is denoted as the higher region. Under the condition that the phase map is adequately sampled in a global sense (see Section 5.1), the regions on either side of a contour can only differ by one cycle. Thus, relative unwrapping is a simple process.

In absolute unwrapping, the phase-offset relationships between all regions are established using the previously established relative offsets. For the first contour, an absolute phase offset is assigned to each of the left and right adjoining regions. The lower of the two is assigned zero cycles while the higher is assigned one cycle. The remainder of the contours is searched for matching left or right regions. If a match is found, then the absolute phase offset is assigned. Using the known relative phase offset, the absolute number of cycles for the other region is then assigned. Once all regions have been accounted for, the number of cycles assigned are multiplied by 2π radians and added to the wrapped phase map. At this point, the two-dimensional phase unwrapping is

essentially complete. The advantage of this algorithm is that by approaching the problem on a contour-by-contour basis, the task of two-dimensional phase unwrapping is collapsed into a one-dimensional problem.

The final step is unwrapping of the contours themselves. In a continuous phase map, contours are curves of infinitesimal width. However, contours have a finite width of one pixel in a sampled phase map. Moreover, the peak response of edge operators can be displaced in noisy imagery (Canny, 1986). Thus, the decision as to which region the contour node pixels should be assigned is not always clear.

The approach taken here is based upon the unwrapped phase of the regions adjacent to the contours. To unwrap a contour-related pixel, only two candidate number of cycles exist: the number cycles of the left and right regions. The following two quantities are calculated for each contour node using the neighbouring region pixels

$$\sum \left| \left(\phi_d + 2\pi N_L \right) - \phi_c \right| \tag{5.6a}$$

$$\sum \left| \left(\phi_d + 2\pi N_R \right) - \phi_c \right| , \tag{5.6b}$$

where

ϕ_d is the wrapped (discontinuous) phase at the contour node.

N_L, N_R are the candidate number of cycles for the left and right regions, respectively.

ϕ_c is the unwrapped (continuous) phase of a neighbouring region pixel.

$\sum$ represents summation over the neighbouring region pixels.

The candidate number of cycles for which the sum of absolute values of differences is least is chosen.

5.3 Chapter Summary

A new algorithm for two-dimensional phase unwrapping has been proposed that is based upon a set of topological rules governing the relationships between various entities in a phase map. Unlike existing approaches, this method treats the unwrapping task as a two-dimensional problem rather than a series of independent, one-dimensional tasks. The various steps of the algorithm are depicted by the flowchart in Figure 5.5.

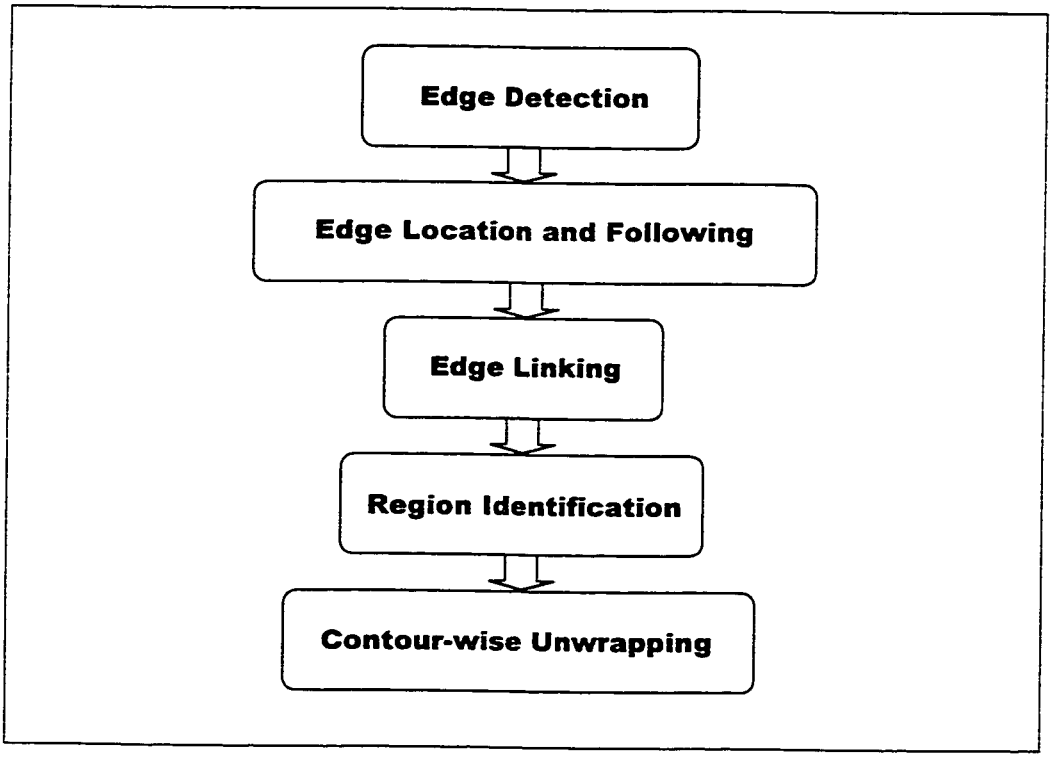


Figure 5.5. Two-Dimensional Phase Unwrapping Algorithm.

CHAPTER 6

TESTING AND RESULTS

6.1 Fringe Extraction

Several real holographic interferometry images have been utilised for testing of the fringe maxima extraction algorithm described in Chapter Four. A simulated image has also been processed. Detailed description of the data as well as the results and analysis of each processing step are presented in the following sub-sections.

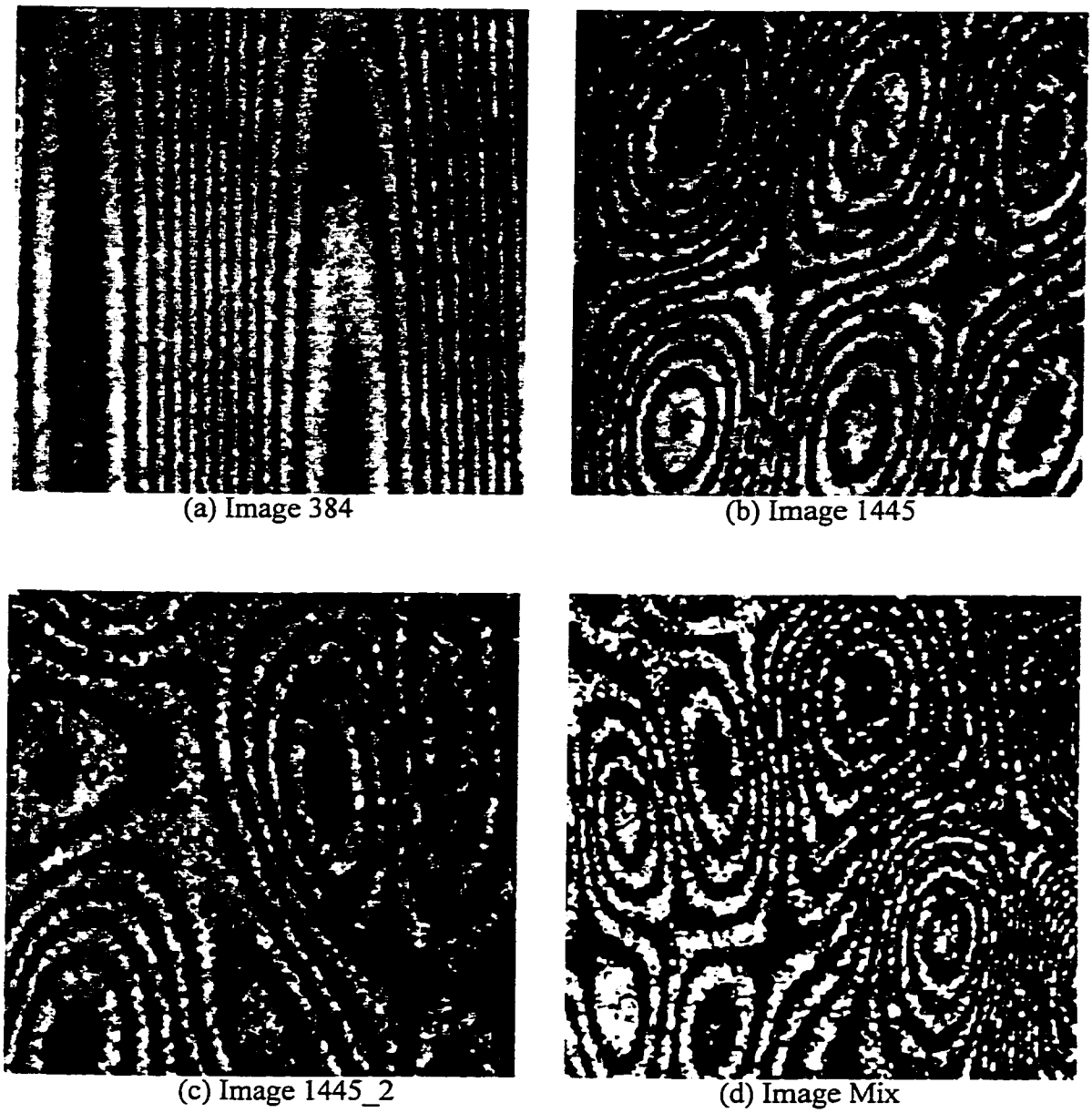
6.1.1 Data Description

Trans-Canada Pipelines, Limited (TCPL), provided several transparency prints of holographic interferograms acquired by Holographics, Inc. for this research. The imagery was originally acquired as part of a preliminary study into vibration measurement of large diameter pipes. Set in a laboratory environment, a barrel was mechanically excited at several different frequencies to simulate a section of pipe vibrating under various operating conditions. Three images were utilised for this research, the first with 384-Hertz excitation, and the second at 1445 Hertz. The third image was captured with the barrel stimulated with a mix of three different frequencies. The laser wavelength of the interferometer was approximately 700 nanometres.

The images provided were produced with a laser printer by TCPL. Thus, they had been previously scanned from the original holographic image prints. The significance of the scanning and, particularly, the printing processes will be discussed shortly. The author using a Hewlett-Packard ScanJet 6100c-flatbed scanner subsequently scanned the prints provided. A spatial resolution of 300 dots-per-inch (dpi) in each dimension was selected. This frequency corresponds to a sampling interval, Δs, of approximately 85 micrometres. This choice was motivated by two factors: storage requirements and fringe pattern frequency. The greatest possible resolution available with the aforementioned hardware was 600 dpi. Imagery scanned at this rate was found to be too large (in terms of storage requirements) to be of practical use. Imagery was also acquired at lower resolutions, but discounted due to poor image quality as it was found that fringe patterns in areas where the displacement field gradient was relatively large were not adequately reproduced. That is, closely spaced fringes were not as easily distinguishable at lower resolutions as they were at 300 dpi. Thus, 300 dpi was empirically found to be the scanning resolution for which the trade-off between storage and over-sampling was best. As for radiometric resolution, 8-bit quantization (256 intensity values) was used.

For easier data handling during the course of the testing, smaller sub-images were extracted from the scanned images. While the full images consisted of approximately 1000 x 3000 samples, the individual working images were 512 x 512. The four test sub-images utilised here are shown in Figure 6.1. The image of Figure 6.1a, denoted as image 384, was extracted from the image of the barrel when shaken at 384 Hertz. Figures 6.1b and 6.1c, denoted as 1445 and 1445_2, respectively, were taken from

different locations of the 1445-Hertz excitation image. Finally, Figure 6.1d, denoted as Mix, corresponds to a portion of the three-frequency mix image.



(a) Image 384 (b) Image 1445

(c) Image 1445_2 (d) Image Mix

Figure 6.1. Working Images for the Fringe Extraction Algorithm Testing.

6.1.2 Pre-filtering

Preliminary spatial and spatial-frequency domain analyses of the scanned images revealed high frequency noise sources in both dimensions. The noise was dominant in the vertical direction. To illustrate, Figure 6.2 is a plot of intensity values along the middle column (number 255) of image 1445. Note the high frequency variations in intensity superimposed on the sinusoidal pattern.

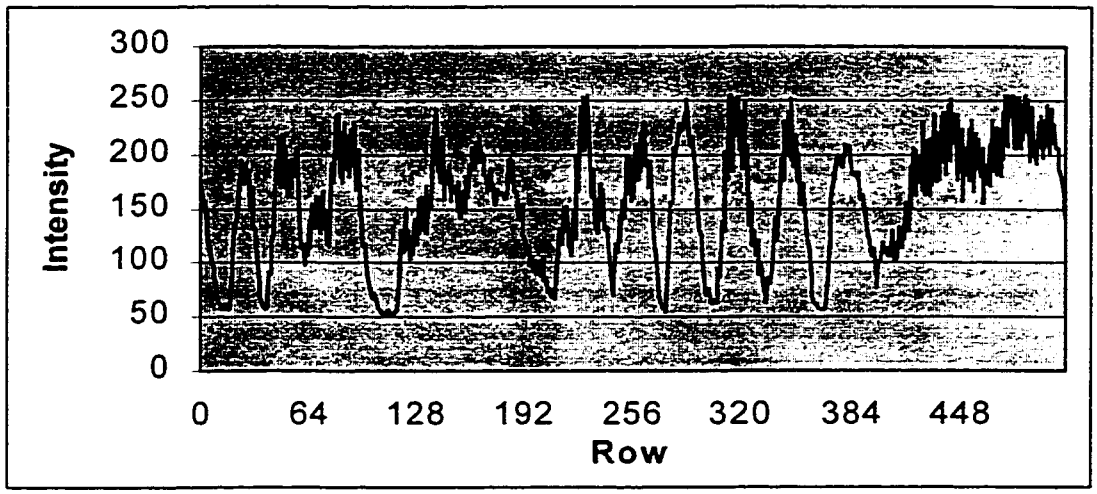


Figure 6.2. Intensity Cross Section along Central Column of Image 1445 Before Filtering.

This noise source is also clearly visible in the spectrum of image 1445. Figure 6.3 is the log magnitude spectrum of image 1445, weighted with a rectangular Hanning window. While most of the useful image content is concentrated about the frequency origin at the image centre, the noise appears at the image extremities along the both central row and central column.

Figure 6.3. Log Magnitude Spectrum of Image 1445 (Figure 6.1b).

After inspection of the transparencies with a magnifying glass, it is believed that the laser printing was the source of this noise. The four pulses containing the noise were removed by multiplying the (non-windowed) spectrum with a pair of symmetric, Gaussian band-stop filters, each of the form

$$H(u,v) = 1 - e^{-\frac{1}{2}\left(\frac{(u-u_0)^2}{\sigma_u^2} + \frac{(v-v_0)^2}{\sigma_v^2}\right)} - e^{-\frac{1}{2}\left(\frac{(u-u_0)^2}{\sigma_u^2} + \frac{(v+v_0)^2}{\sigma_v^2}\right)}, \qquad (6.1)$$

where

u_0, v_0 are the central frequencies in the row and column directions, respectively,

σ_u, σ_v are the standard deviations (bandwidth estimates) in the row and column directions, respectively.

An exemplary surface plot of Equation 6.1 is given in Figure 6.4. The central frequencies and bandwidths were determined by inspection of the magnitude spectra. The filtering

proved to be effective at removing the noise, as indicated by the smoother cross-section shown in Figure 6.5.

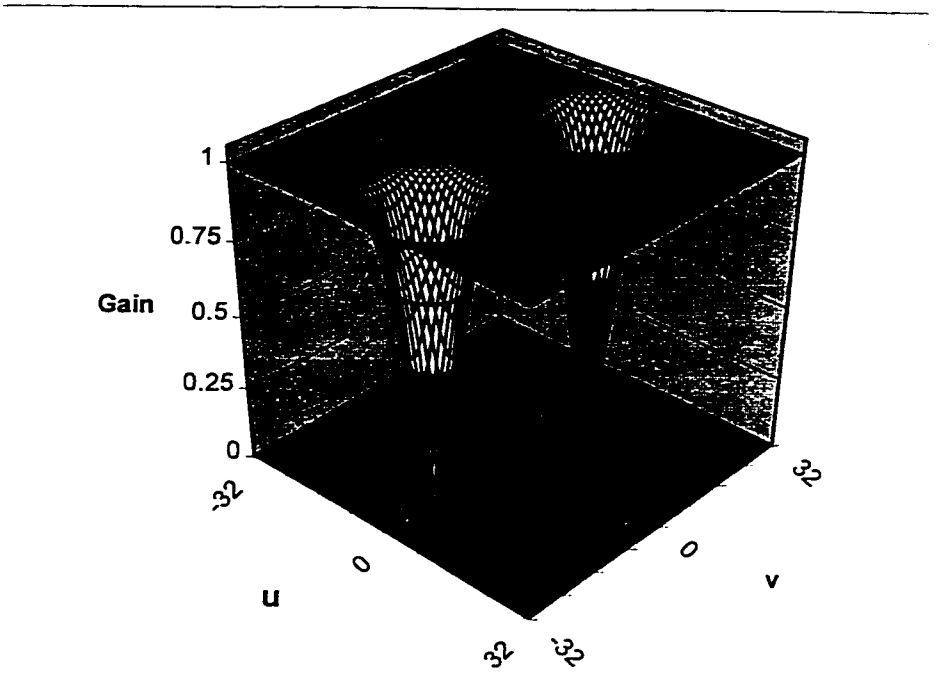


Figure 6.4. Symmetric, Gaussian Band-Stop Filters.

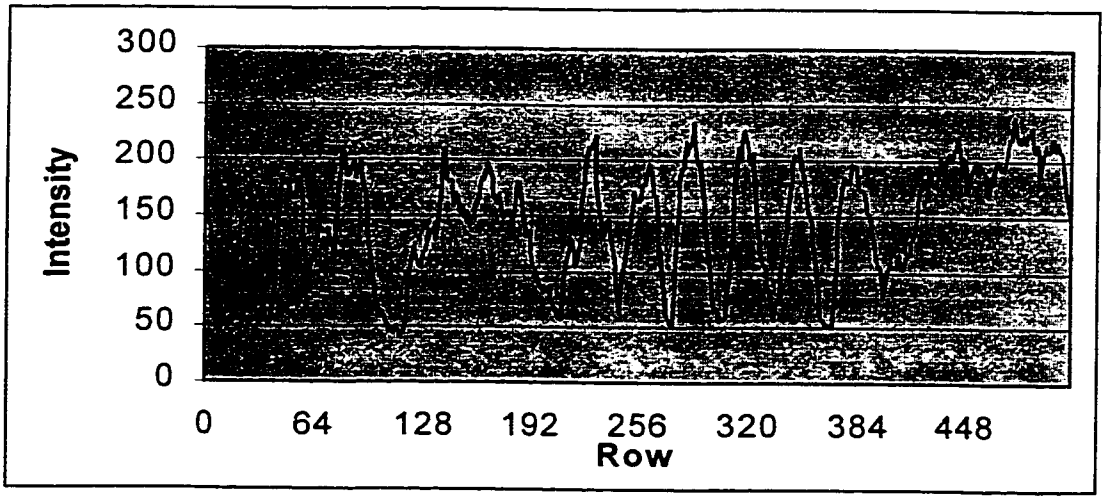
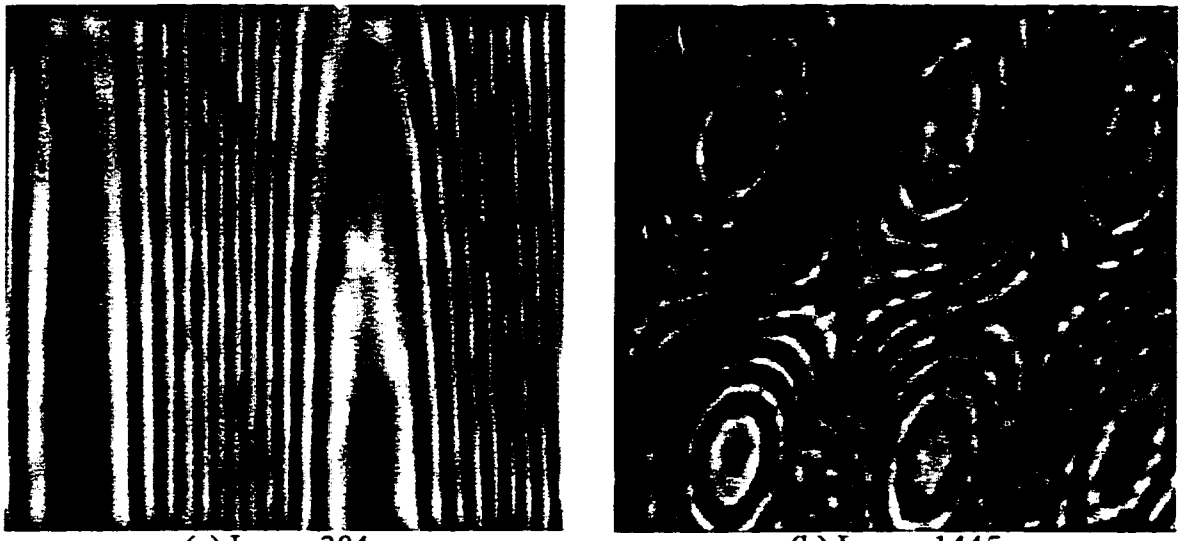


Figure 6.5. Intensity Cross Section along Central Column of Image 1445 After Filtering.

6.1.3 Homomorphic Filtering

The first step of the fringe maxima extraction algorithm described in Chapter Four is speckle noise reduction via homomorphic filtering. The speckle is clearly visible in all four images of Figure 6.1, but Figures 6.1c and 6.1d are particularly contaminated. Due to the spatial correlation of speckle noise, the distortion appears to be the most severe in areas with narrow fringe spacing. This is poses a serious problem in the task of extracting the fringe maxima, as the speckle creates false intensity valleys and ridges.

After visual determination of appropriate bandwidths, all images were homomorphically filtered with fourth-order Butterworth functions (see Equation 4.2). Experiments were conducted with various filter types, including Gaussian and lower order Butterworth functions. The fourth-order Butterworth was found to perform the best of those tested. The images produced by this operation exhibit overall improved appearance, as shown in Figure 6.6, in terms of reduced speckle. However, due to the attenuation of high-frequency details, they are blurred to some degree. The resulting distortion is most severe in the areas of narrow fringe spacing. In particular, note the loss of fringe detail in the lower right-hand corner of Figure 6.6d.



(a) Image 384 (b) Image 1445

(c) Image 1445_2 (d) Image Mix

Figure 6.6. Homomorphically Filtered Working Images.

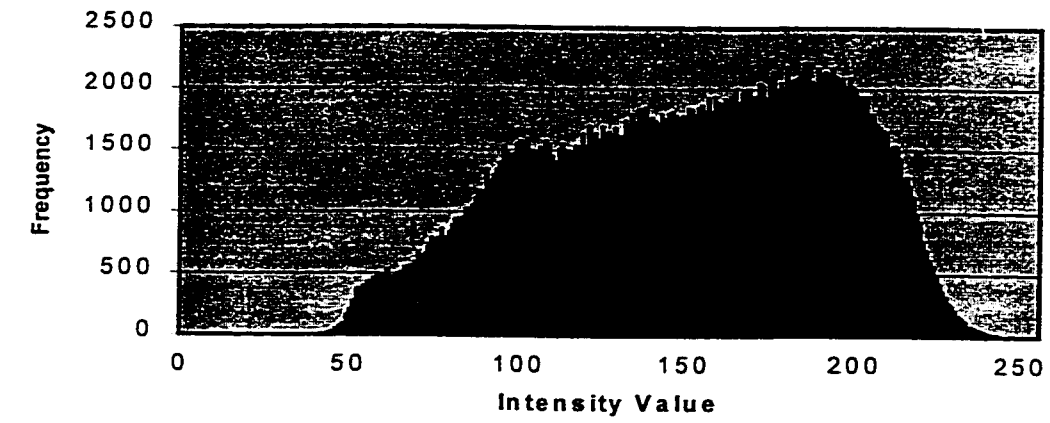
Figure 6.6a, although slightly blurred, was dramatically improved by the filtering. Note the distortions at the edges caused by the periodic extensions of the image contributing to the filter response. No zero padding of the imagery was performed prior

to the 2-D FFT and subsequent filtering. The images in Figures 6.1b and 6.1c are similarly improved, but with blurring in the areas of narrow fringe spacing. False intensity maxima have been created in some locations with large gradients due to the blurring. These distortions will be addressed in subsequent subsections.

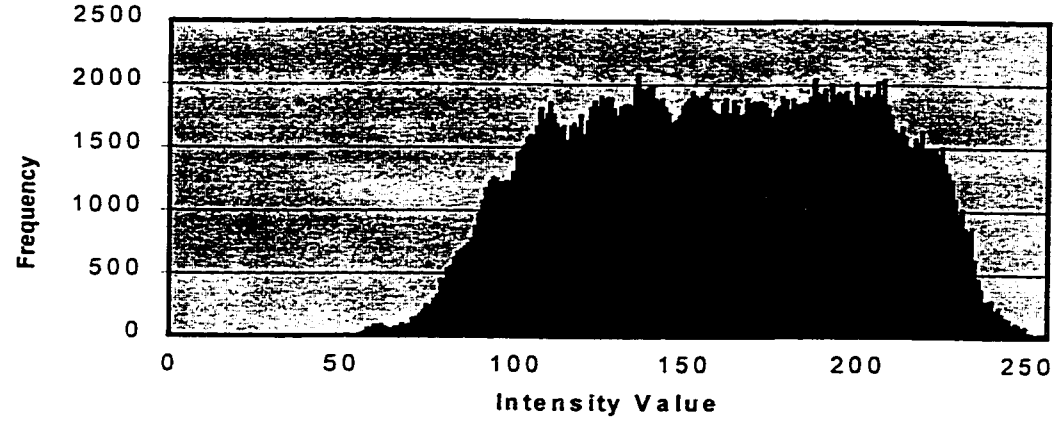
6.1.4 Thresholding

The need for decomposing the homomorphically-filtered imagery into smaller regions prior to thresholding is best understood by analysing Figure 6.4. This intensity profile exhibits non-linear modulation depth, as indicated by the varying height of intensity maxima. This makes global thresholding inappropriate. Square regions of 64 x 64 pixels were found to be most effective in terms of capturing less intense fringe maxima while maintaining separation between stronger fringe maxima.

The problem of choosing an appropriate threshold can be better understood by analysing the histograms of image 384 presented in Figure 6.6. Figure 6.7a represents the global histogram prior to homomorphic filtering, and the post-filtering histogram is given in Figure 6.7b. Note that both histograms possess a unimodal shape.



(a) Histogram Prior to Homomorphic Filtering



(b) Histogram After Homomorphic Filtering

Figure 6.7. Histograms of Image 384.

Figure 6.8 presents both global and local thresholding results for image 1445_2. Areas of significance in Figure 6.8a (globally thresholded image) are the lower left portion where adjacent fringes are joined with bridges and at the top of the image where the fringes are broken by several gaps. In the locally thresholded image (Figure 6.8b), the bridges between adjacent contours in the lower left do not exist. Furthermore, the

fringes at the top of this image are also more continuous (they have fewer gaps). Note, however, the distortion in both images where the fringe spacing is narrow. This is clearly visible between the two elongated, closed loop fringes in the upper right quadrant of the images.

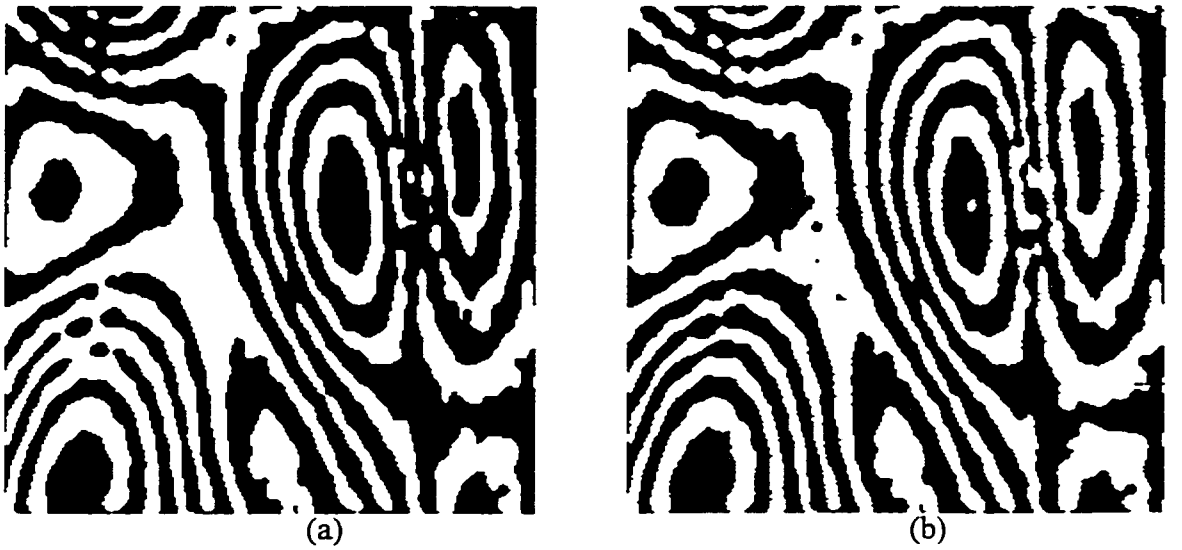
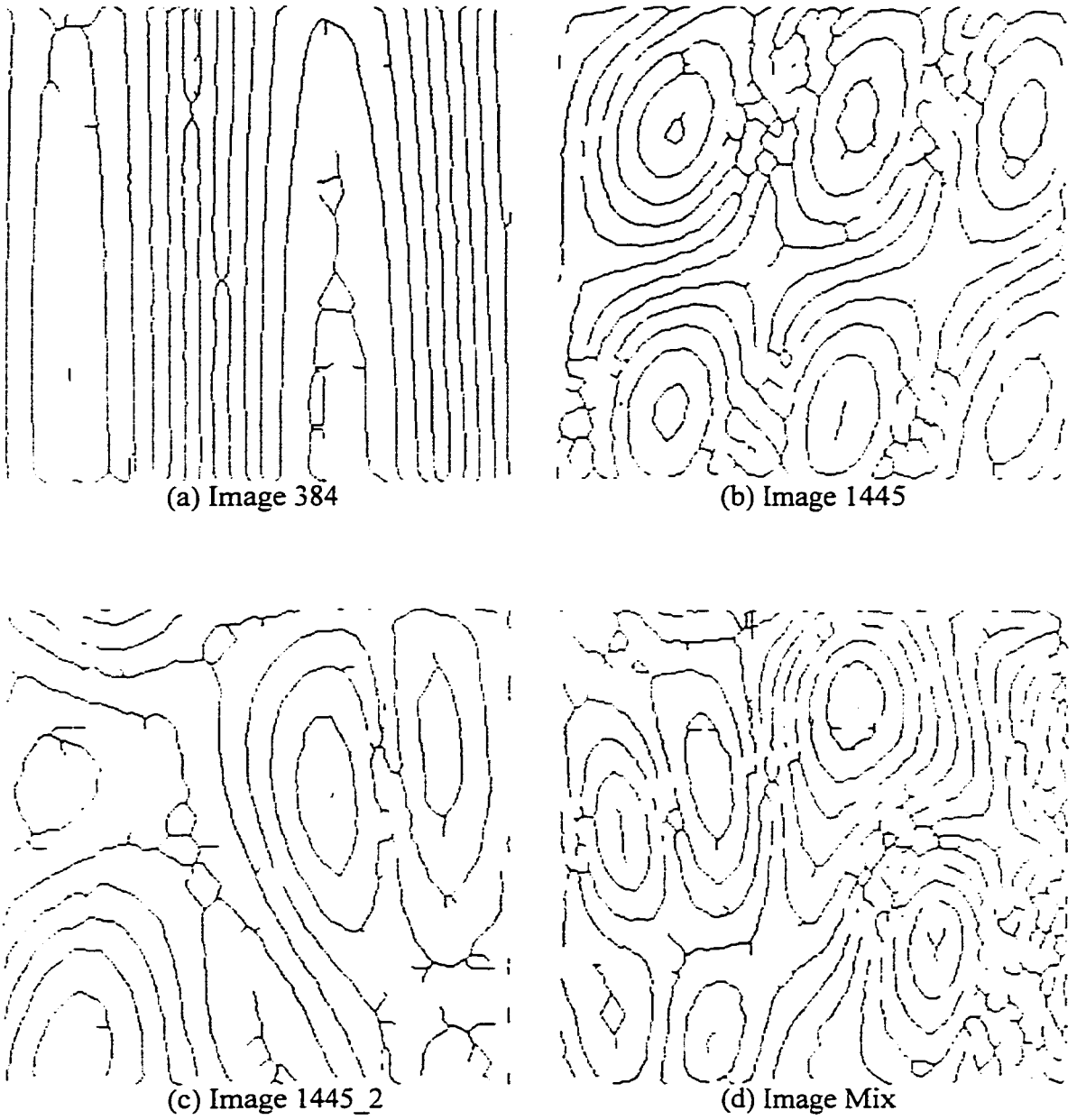


Figure 6.8. Global and Local Thresholding Results (Image 1445_2).

6.1.5 Thinning

As previously mentioned the binary morphological thinning operation reduces nonzero sets in an image to single pixel-wide entities. The thinned fringe maxima are given in Figure 6.9. Spur, bridge and gap artefacts as well as false contours are present in all four images. The images corresponding to 1445 and Mix in particular are highly contaminated with artefacts. In fact, a great deal of semiautomatic editing was required in the right-hand portion of the latter image due to the distortion. As expected, the

occurrence of artefacts is greater in the areas of narrow fringe spacing, where the blurring is most severe.



(a) Image 384 (b) Image 1445

(c) Image 1445_2 (d) Image Mix

Figure 6.9. Thinned Fringe Maxima (Contrast Reversed).

As previously mentioned, the process of homomorphic filtering contributes to the problem of contour artefacts due to its blurring effect. However, substantial improvement in the quality of thinned contours is realised by employing the filtering. As an example, Figure 6.10 shows the thinned fringe maxima from image 1445 without application of the homomorphic filtering. The presence of numerous additional spur and bridge artefacts is quite evident when this image is compared with Figure 6.9b. Although the true fringe maxima in the areas of narrow fringe spacing are less distorted in Figure 6.10 than in Figure 6.9b, the spurs and bridges still create the need for semiautomatic editing. The goal of the homomorphic filtering then is to remove as much noise as possible such that the thinned contour map is as clean as possible for the extraction and editing stages.

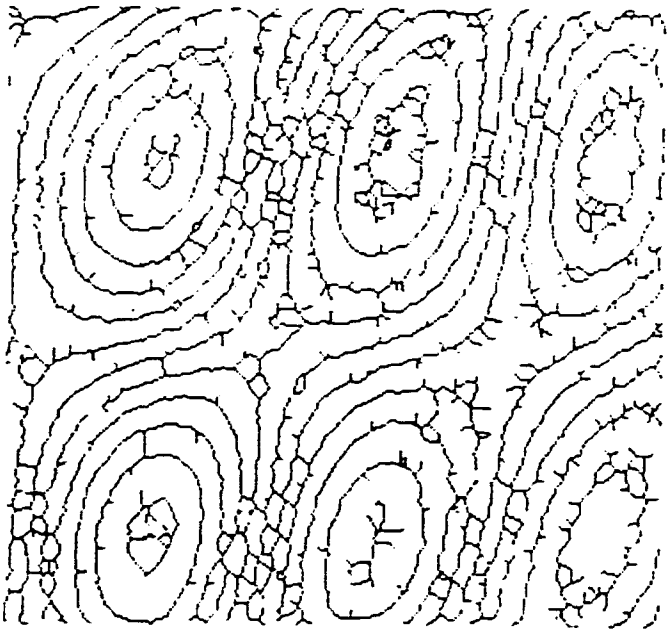


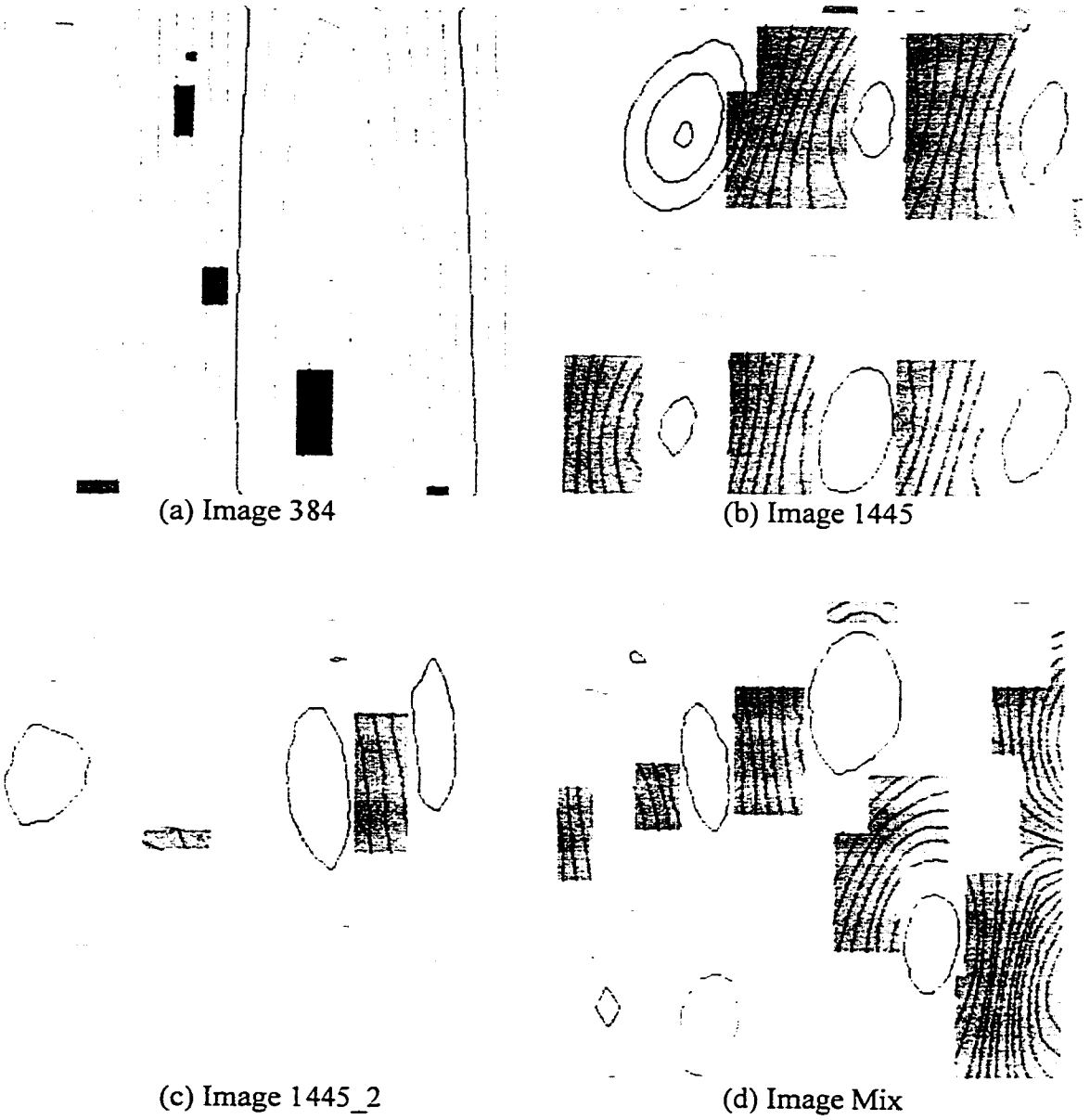
Figure 6.10. Thinned Fringe Maxima of Image 1445 Without Homomorphic Filtering (Contrast Reversed).

6.1.6 Automatic and Semiautomatic Editing

Plotted in Figure 6.11 are the extracted contours from each image after automated editing (spur, bridge and gap removal) and semiautomatic contour interpolation. The shaded areas in these figures indicate the windows in which interpolation was performed. As can be gathered from Figure 6.9, little editing of extracted contours was required for images 384 and 1445_2. The homomorphic filtering of these two images was very successful at reducing the speckle, thereby yielding cleaner contours. Nevertheless, some of the semiautomatic procedures detailed in Chapter Four were required. For image 384 (Figure 6.11a), only a few small window areas were required for which simple linear interpolation was sufficient. For image 1445_2 (Figure 6.11c), only three window areas were necessary, one for linear interpolation (near the image centre) and two for concentric contour matching. Note the presence of ridge contours in both images.

Images 1445 and Mix required considerably more semiautomatic editing. Concentric contour matching was successfully applied in a few locations of image 1445 (Figure 6.11b). This was due to the smoothness of the upper left, lower left and lower middle closed contours. However, the matching procedure could only be applied once or twice per contour due to the amplification of noise in the inner contour. The remainder of the windowed gaps was bridged with quadratic interpolation. Image Mix (Figure 6.11d), required considerably more semiautomatic editing. While concentric contour matching was successfully applied in several locations, including the left and central portions of the image, a considerable amount of quadratic interpolation was required. As with image 1445, the contour matching was restricted to one or two applications per

closed contour. The distorted right-hand side of image Mix is almost completely interpolated using parabolas. This image also includes a few degenerate contours (ridges and a peak).



(a) Image 384 (b) Image 1445

(c) Image 1445_2 (d) Image Mix

Figure 6.11. Edited Fringe Maxima Contours.

6.1.7 Efficiency and Accuracy of Fringe Maxima Extraction

In order to evaluate the effectiveness of the fringe maxima extraction algorithm described in Chapter Four, two criteria were used: efficiency and accuracy. Efficiency is an internal quantification of the algorithm's effectiveness and is measured by the percentage of contour nodes extracted automatically. The evaluation of accuracy is based upon both theoretical and empirical measures of planimetric contour positioning and displacement estimation. Unfortunately, neither theoretical computations nor independent measurements (i.e., from accelerometers) were available for accuracy assessment. Thus, only the proposed algorithms and not HI itself could be assessed.

The efficiency of contour extraction for each image is summarised in Table 6.1. The total number of contour nodes for each image is broken down into the number and percentage extracted both automatically and using the semiautomatic methods. For images 384 and 1445_2, the efficiency was quite high at 96.4% and 91.2%, respectively. Images 1445 and Mix, though, had considerably poorer efficiency measures at 54.4% and 51.7%, respectively. The lower ratings are due to the severity of distortions and artefacts in these images. However, a poor efficiency rating is deceiving since the number of nodes extracted in semiautomatic mode is not necessarily indicative of the amount of manual work involved. For example, linear interpolation only requires that the user identify two points that define the window boundaries. Thus, the number of points that the operator must delineate is considerably less than the figures indicated in the columns of Table 6.1 that refer to semiautomatic contour node extraction.

Image	Number of Nodes Extracted			Percent Extracted	
	Total	Automatic	Semiautomatic	Automatic	Semiautomatic
384	11083	10796	287	96.4	2.6
1445	11489	6246	5243	54.4	45.6
1445_2	6699	6110	589	91.2	8.8
Mix	13168	6941	6227	52.7	46.3

Table 6.1. Fringe Maxima Extraction Efficiency.

The quantification of contour accuracy requires detailed analysis of all error sources. If the goal of fringe maxima extraction is simply to determine peak antinode deflection, then the absolute position of each contour is not critical. However, one should expect that an extracted contour lie completely within its respective bright fringe. Thus, in assessing the accuracy of antinode displacement, planimetric accuracy need not be accounted for. The noise sources affecting derived displacements are given in the following list.

- Speckle.
- Film grain density.
- Sampling.
- Quantization.
- Electronic.
- Homomorphic filtering.

All of these noise sources introduce distortion into the intensity image, from which the contours and, ultimately, deformations are derived.

If the goal of the contour extraction is to estimate mode shapes and/or the exact positions of peak antinode displacement, then planimetric contour accuracy is a significant factor. In addition to the error sources given above, errors due to the thresholding and morphological thinning processes must also be included.

6.1.8 Quantification of Noise Sources

Many of the error sources listed in the previous subsection are image dependent. In order to facilitate simplified quantification, a worst-case scenario image was utilised. The shortest spatial wavelength fringe observed in all of the four test images was 6 samples (approximately 500 μm). This sinusoid is used for some of the analyses below. Full modulation depth, constant background intensity and 8-bit quantization have been assumed. The worst-case fringe pattern has been assumed to parallel the image x-axis and is given by

$$i(x,y) = 127.5(1 + \cos(f_m x + \theta)) \ , \qquad (6.2)$$

where

f_m is the worst-case spatial frequency of 0.002 μm^{-1},

θ is a (random) spatial phase shift.

6.1.8.1 Speckle

As detailed in Subsection 2.4.1, laser speckle is a multiplicative noise source that depends upon surface roughness. The random variation in phase, ϕ, due to this roughness and the use of coherent radiation is modelled as having a uniform density on $[-\pi,\pi]$. The standard deviation of this density is $\sigma_\phi \pm \pi(3)^{-1}$. Using Equations 3.2, 3.3 and 3.4, the proportional standard deviation of the deformation due to speckle, σ_{sp}, is

$$\sigma_{sp} = \pm 58 \quad \text{nm} \ . \tag{6.3}$$

This represents approximately ±17% of HI resolution and clearly indicates the need for homomorphic filtering. It should be noted that the auto-correlation function given by Equation 2.38 is not applicable to the imagery used here due to the cylindrical surface shape of the barrel.

6.1.8.2 Film Grain Density

The Nutting model relates grain size and number and scanning aperture size to the optical density of a photographic emulsion (Castleman, 1996). The basis for the model is the hypothesis that a photographic emulsion is comprised of composite layers of film grains, each one-grain in thickness. The empirically derived model for density is given by (Castleman, 1996)

$$D = 0.43\frac{na}{A} \ , \tag{6.4}$$

where

D is the optical density,

n is the number of grains within the scanning aperture.

a is the cross-sectional area of a grain.

A is the diameter of a circular scanning aperture.

If the grain size is much less than the aperture, then n is Poisson distributed (Castleman, 1996). Consequently, the standard deviation in optical density is a function of the average number of grains within the aperture, since the mean and variance of the Poisson density are equal.

$$\sigma_D = \pm 0.43 \frac{a}{A} \sqrt{E\{n\}} \tag{6.5}$$

Unfortunately, the pertinent details of the photographic emulsion used for the test imagery were not made available to the author. In an attempt to realistically quantify film grain noise, a mean film grain diameter of 40 μm has been assumed. This corresponds to the upper limit of grain size quoted for fine-grained emulsion holoplates produced by the SLAVICH Company and distributed by the GEOLA Company of Lithuania. This information was obtained from GEOLA's world-wide website (www.geola.com). For the scanning aperture, the area of the 85μm x 85μm scanner sampling aperture was used to calculate the diameter of a circular aperture with equivalent area. This diameter was then reduced by a factor of 6.55 to account for the enlargement process from 35mm film to the size of the fringe image prints. The final diameter was 12.7μm. The mean number of grains was calculated to be 280 by assuming they form a rectangular lattice within the reduced rectangular sampling aperture. Using

these values and Equation 6.5, the standard deviation of optical density due to film grain noise was calculated to be ±0.0060. Using the relation between density and transmittance

$$D = -\log T \ , \tag{6.6}$$

the standard deviation of transmittance can be estimated via variance propagation:

$$\sigma_T = \pm\sqrt{\left(\frac{dT}{dD}\right)^2 \sigma_D^2}$$

$$= \pm \ln 10 \cdot e^{-D \cdot \ln 10} \sigma_D \tag{6.7}$$

From the negative exponential relationship of Equation 6.7, it can be seen that transmittance error decreases as density increases. In order to transform numbers calculated via Equation 6.7 into a standard error in displacement, the H-D curve for the film along with exposure time is required. Since this information was not available, a linear relation between transmittance and quantized intensity was assumed. The assumption is that a transmittance of 0.1 (density of 1.0) corresponds to an intensity of 126.5 (the mean value of the assumed worst-case sinusoid). Table 6.2 is a presentation of standard errors in displacement corresponding to different values of density. Using a density of 0.1 for a worst case scenario, the standard deviation in displacement due to film grain noise is

$$\sigma_{fg} = \pm 19.2 \quad \text{nm} \, . \tag{6.8}$$

This represents ±5.5% of the HI measurement resolution of 350nm.

D	σ_T	$\sigma_{\Delta Z}$ (nm)
0.1	±0.0110	±19.2
0.5	±0.0044	±6.7
1.0	±0.0014	±2.4
2.0	±0.00014	±0.24

Table 6.2. Standard Deviation in Displacement for Different Values of Density.

6.1.8.3 Sampling

The sampled version of Equation 6.2 using a finite-area aperture, as is found in charge-coupled devices, is given by Equation 6.9 (Pratt, 1976).

$$i_s(m\Delta s, n\Delta s) = \frac{\int_{m-\frac{\Delta s}{2}}^{m+\frac{\Delta s}{2}} \int_{n-\frac{\Delta s}{2}}^{n+\frac{\Delta s}{2}} i(x,y)\,dxdy}{\int_{m-\frac{\Delta s}{2}}^{m+\frac{\Delta s}{2}} \int_{n-\frac{\Delta s}{2}}^{n+\frac{\Delta s}{2}} dxdy} = \frac{1}{\Delta s^2} \int_{m-\frac{\Delta s}{2}}^{m+\frac{\Delta s}{2}} \int_{n-\frac{\Delta s}{2}}^{n+\frac{\Delta s}{2}} i(x,y)\,dxdy \tag{6.9}$$

This equation indicates that a discrete sample of the intensity image is equal to the average intensity over the sampling aperture normalised by the aperture area. For the sake of simplicity, the discrete indices m and n are both set to zero. Then, after substitution of Equation 6.2 into Equation 6.9 and performing some algebraic manipulation, the sampled intensity of the worst-case fringe at $m = n = 0$ is

$$i_s(0,0) = 127.5\left(1 + \frac{3}{\pi}\cos\theta\right). \tag{6.10}$$

Thus, the sampled intensity of the worst-case fringe is given as a function of its spatial phase (position). The radiometric sampling error, e_r, is defined as the difference between the incident and sampled versions of the image, as shown in Equation 6.11.

$$e_r(m,n) = i(m,n) - i_s(m,n) \tag{6.11}$$

This error is caused by the difference between the actual average values of an image at a given location and its magnitude governed by the spatial phase shift.

In order to quantify the effect of a random phase shift on the intensity and, ultimately, displacement, knowledge of the probability density function (pdf) is required. If θ has a uniform density on $[-\pi, \pi]$, then all phase shifts on this interval have an equal probability of occurrence of $(2\pi)^{-1}$. Given the pdf of θ, $f(\theta)$ and its functional relationship with e_r, $g(\theta)$, the pdf of e_r, $f(e_r)$, can be determined by (Papoulis, 1991)

$$f(e_r) = \sum_{i=1}^{n} \frac{f(\theta_i)}{\left|\left.\frac{dg(\theta)}{d\theta}\right|_{\theta=\theta_i}\right|}, \tag{6.12}$$

where

θ_i is the i^{th} solution to $e_r = g(\theta)$,

n is the total number of possible solutions to $e_r = g(\theta)$.

Since $f(\theta)$ is only nonzero on $[-\pi, \pi]$, then only two solutions to $e_r = g(\theta)$ exist. It can be shown that the pdf e_r is

$$f(e_r) = \frac{1}{\pi\sqrt{a^2 - e_r^2}}, \tag{6.13}$$

where

$$a = 127.5\left(1 - \frac{3}{\pi}\right). \tag{6.14}$$

This pdf is concave-upward and symmetric about zero with a value of $(a\pi)^{-1}$ at the origin. It has vertical asymptotes at $\pm a$ and a standard deviation of $\pm 2^{-0.5} a \approx \pm 4.06$. Using Equations 3.2, 3.3 and 3.4, the corresponding standard deviation in deformation due to a random phase shift (due to sampling), σ_{sa}, is

$$\sigma_{sa} = \pm 5.6 \quad \text{nm}. \tag{6.15}$$

This standard error represents only about ±1.6% of the HI measurement resolution.

6.1.8.4 Quantization

Quantization noise due to rounding has been modelled by Oppenheim and Schafer (1989) as additive white noise with a uniform density. This pdf has a value of one on $(-0.5, 0.5)$ and a standard deviation of ±0.289. The corresponding error in deformation due to quantization, σ_q, is

$$\sigma_q = \pm 0.40 \quad \text{nm}. \tag{6.17}$$

This figure represents only about ±0.1% of the measurement resolution.

The combined effect of speckle, film grain, sampling and quantization noises for the worst case situation outlined earlier in this chapter is given by

$$\sigma = \pm\sqrt{\sigma_{sp}^2 + \sigma_{fg}^2 + \sigma_{sa}^2 + \sigma_q^2} \tag{6.17}$$

$$= \pm 61.4 \quad \text{nm}$$

Clearly the speckle noise is dominant, as it is at least an order of magnitude greater than the other sources except for the film grain density.

6.1.8.5 Other Noise Sources

Since the analytical form of the deformation surfaces (and their spectra) within the test images was not known, the amount of noise added by the processes of homomorphic filtering was not known. The level of electronic noise also was not known.

6.1.9 Accuracy Quantification

As previously mentioned, a rigorous quantification of deformation field measurement accuracy was not possible for this study. However, an empirical study has been conducted. The fringe maxima contours were manually delineated for images 384 and 1445_2. These particular images were selected because they required the least amount of contour interpolation and contained the least complex fringe patterns. The difference between the two sets of contours and their respective automatically extracted counterparts was taken as the basis for evaluation. The manually digitised images were used as for reference. The root mean square error (RMSE) of the co-ordinate differences was calculated for three different areas within each image. Each area consisted of approximately 100 x 100 samples and was randomly selected from portions of the imagery that were not interpolated and did not contain large differences. Such

differences manifested themselves at very broad fringes, such as near the centre of image 384, and at saddle points, as in image 1445_2. The image borders were also avoided in order to discount biases due to edge effects. The contours used from image 384 were fairly straight with a few minor bends. The contours of image 1445_2, however, were much more curved and complex. The statistics are summarised in Table 6.3

Analysis of these results indicates that the positional accuracy of contour positioning in image 384 was homogeneous throughout the image at about ±1 pixel. Large discrepancies, which occurred at broad fringes, are due to the combined effect of the error sources previously described and personal digitizing bias. The inferior RMSE figures and lack of homogeneity of image 1445_2 can be attributed to both the greater complexity of its fringe patterns and the broader fringes. Although this evaluation does not represent an absolute but only relative test of contour positional accuracy, it does represent an interesting consistency check between the two methods, since manual digitization is frequently used for contour extraction. Although large discrepancies were encountered at broad fringes, the effect on mode shape is not as severe due to the shallow gradient in such locations.

	Image 384		Image 1445_2	
	Number of Nodes	RMSE (pixels)	Number of Nodes	RMSE (pixels)
Area 1	557	±1.1	369	±3.2
Area 2	624	±1.1	404	±1.9
Area 3	680	±1.0	355	±1.6

Table 6.3. Empirical Fringe Positional Accuracy Statistics.

6.1.10 Simulated Image Contour Extraction

The simulated image of Figure 2.5 was also processed using the fringe maxima extraction algorithm. The extracted contours overlain on the original image are shown in Figure 6.12. Analysis of this figure and comparison with the true contour lines shown in Figure 4.1 reveals biases at the edges. The thinning process has closed the outermost contour that should be broken into four pieces by the image border. Since real imagery is distorted at the borders due to filtering edge effects, this region is not of great concern.

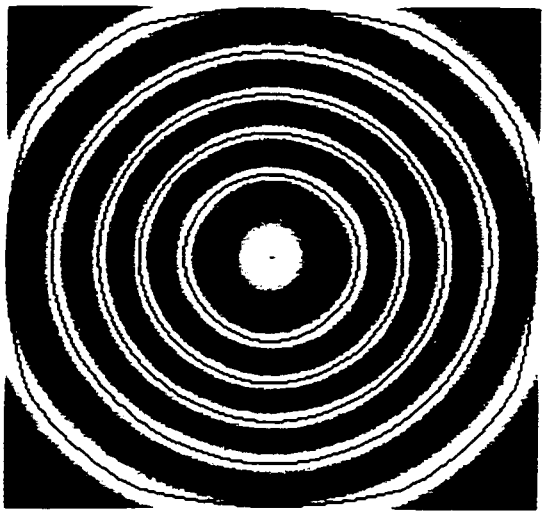


Figure 6.12. Contours Extracted from the Simulated Image of Figure 2.5.

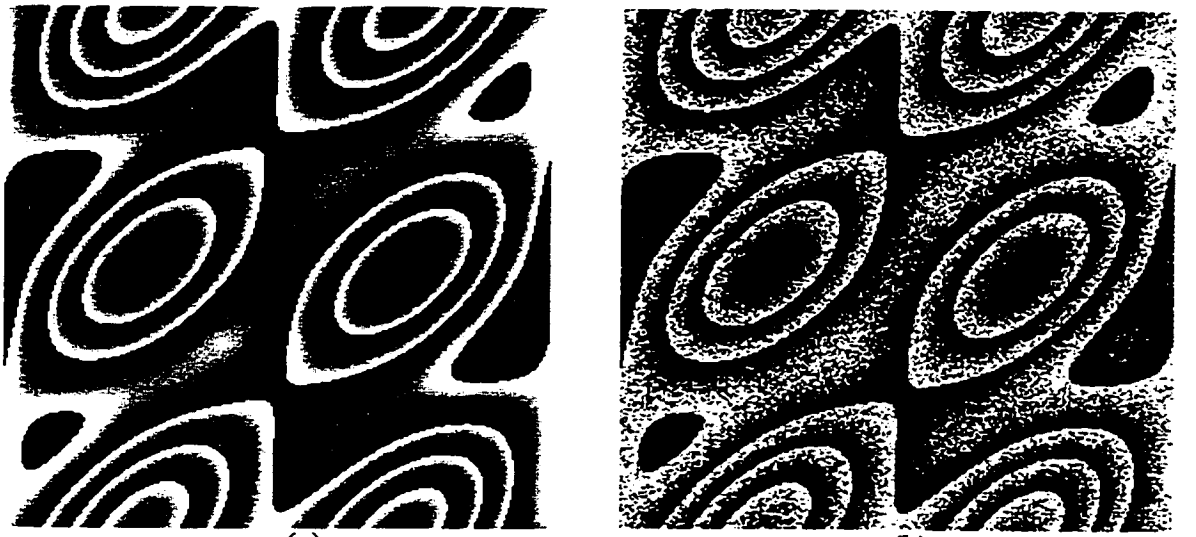
6.2 Phase Unwrapping

Unfortunately, no real off-axis holographic imagery was available during the course of this research for testing the phase unwrapping algorithm described in Chapter Five. Instead, preliminary testing with simulated imagery has been conducted. A detailed description of the data and how it was simulated is presented followed by the processing steps.

6.2.1 Data Description

The simulated phase map used for testing the new two-dimensional phase unwrapping algorithm represents a sinusoidal surface comprised of six terms of a bivariate cosine series. This surface could represent a vibration mode of a thin plate subjected to a periodic mechanical excitation, a situation similar to the vibrating pipe

sections described in Section 6.1.1. The amplitude range of the unwrapped phase surface is approximately 20.8 radians. Uniformly distributed noise (on [-0.8π, 0.8π] radians) was added to the wrapped phase map in an attempt to simulate the effect of speckle. While speckle is modelled as multiplicative noise in intensity, it is additive in phase. Furthermore, although the phase distribution theoretically should be over [-π, π] radians, the noise degradation achieved is quite appreciable. Shown in Figure 6.13a is the simulated, wrapped phase map. The noise-corrupted version is given in Figure 6.13b.



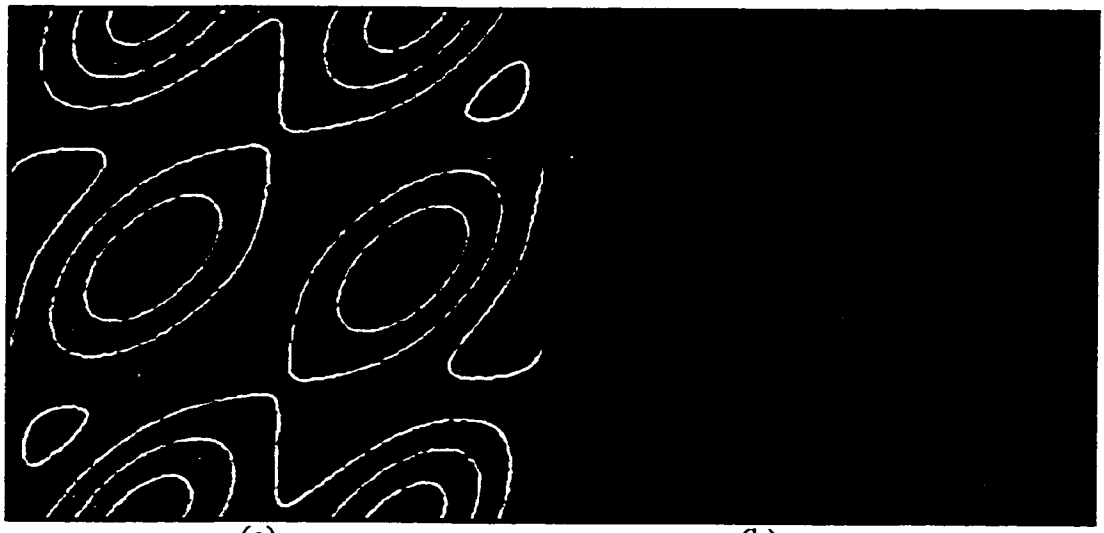
(a) (b)

Figure 6.13. Original and Noisy Simulated Phase Maps.

6.2.2 Data Processing

The first step of the unwrapping algorithm was discontinuity detection. In spite of the noise corruption, the edges of Figure 6.13b are still fairly distinct. The magnitude response images resulting from the application of Equations 5.2a and 5.2b for the first two scales, 1 and 2, are given in Figure 6.14a and 6.14b, respectively. Both images

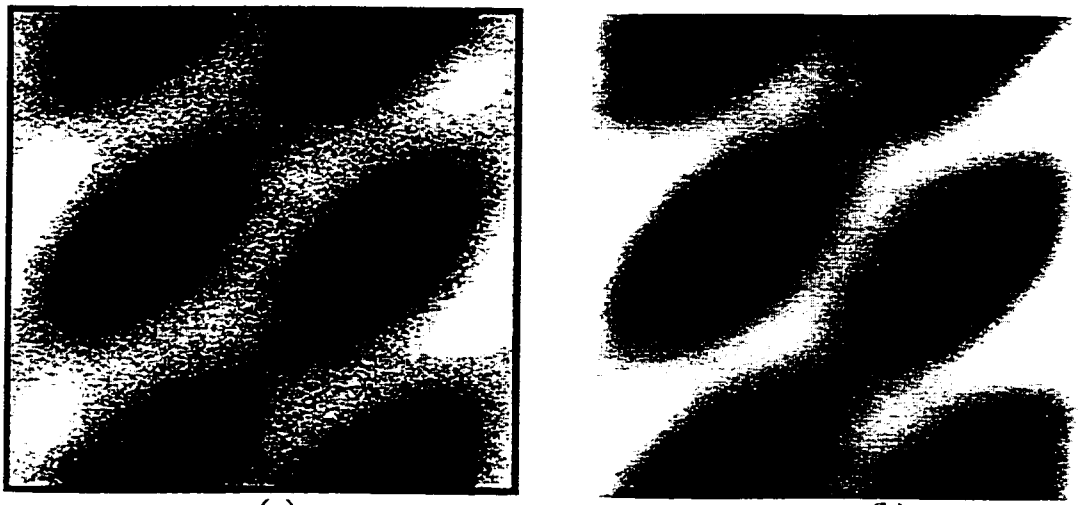
exhibit strong responses at the discontinuity locations. The response at scale 1 appears greater than that of scale 2 because of the scale dependence in the denominator of Equations A.22 and A.23. Note also that the responses of Figure 6.14b are more spread out due to the greater spatial extent of the scale 2 kernel.



(a) (b)

Figure 6.14. Edge Detection Magnitude Responses.

After combination of the two response images via Equation 5.4 and edge response magnitude and direction calculation, the contour following was performed. Subsequent region identification and unwrapping produced the continuous phase map shown in Figure 6.15a. Note that the noise appears to be diminished due to requantization for display purposes. The original, unwrapped phase map is given in Figure 6.15b for comparison.



(a) (b)

Figure 6.15. Noisy and Original Unwrapped Phase Maps.

CHAPTER 7

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

8.1 Summary

Holographic interferometry is an invaluable technique for making high-resolution deformation measurements in many engineering disciplines. Using coherent radiation imaging, HI captures displacement fields as sinusoidal interference (fringe) patterns. Although the means of extracting the information of interest varies according to the imaging geometry, all methods share the common goal of recovering a phase map from a fringe pattern.

For fringe-based analysis, deformations are inferred from contour lines delineating fringe intensity maxima. Extraction techniques include manual contour digitising, which can be extremely time consuming, and automated approaches based upon searches for local intensity maxima. The success of the latter method has been limited to noise-free images exhibiting high fringe contrast.

The automated approach developed in this dissertation was demonstrated to be effective at fringe maxima extraction from real HI imagery. The basis of this method was a topological description of interference fringe imagery outlined in Chapter Four. These

rules provided practical constraints for the editing procedures utilised in the final stage of processing.

The first step of the algorithm was speckle noise reduction via homomorphic filtering. The contour extraction phase consisted of local thresholding followed by morphological processing to thin the segmented fringe maxima. Once extracted via line following, the contours were edited using both automatic and semiautomatic methods. The product of the algorithm was a set of complete contour lines delineating the fringe maxima of an interferogram.

For some applications a more detailed deformation surface is required, necessitating the entire phase map contained within an interferogram. Recovery of this surface entails several digital-filtering operations and, in fact, yields a discontinuous surface. The current crux in processing such imagery is automatic removal of these discontinuities, known as phase unwrapping. Existing techniques to accomplish this essentially treat the phase map as a sequence of independent, one-dimensional signals, each of which is individually unwrapped. The drawback with this approach is that important correlation between neighbouring sample rows (or columns) is ignored during the process of establishing the location of the discontinuities.

An improved algorithm developed in this dissertation exploits this correlation by treating phase map unwrapping as a two-dimensional problem. Preliminary testing using a simulated, noisy phase map proved the algorithm to be successful. Topological rules similar to those of fringe imagery were defined for phase maps in Chapter Five. The algorithm commenced with the detection of discontinuity contour lines using a multi-

scale bank of edge detecting filters. Fusion of the filter responses provided a means to bridge gaps in contour lines caused by noise. After extraction of the contours via line following, it was shown that the actual unwrapping step of the algorithm collapses to a one-dimensional problem.

8.2 Conclusions

From the test results presented in Chapter Six, several conclusions have been drawn concerning the developed methods and these are presented in the following list.

- The critical step of the new algorithm is the homomorphic filtering for speckle noise reduction. A worst-case theoretical study of some of the error sources contributing to the precision of inferred deformations from HI fringe imagery indicated that the speckle was the dominant noise source (see Chapter Six). The standard error of this source was found to be approximately seventeen percent of the measurement resolution of one-half of the laser wavelength.
- The speckle reduction via homomorphic filtering was demonstrated to be largely successful in terms reducing the number of contour line artefacts and, hence, the amount of editing required. However, the overall improvement realised by the filtering came at the expense of image blurring, which proved to be most problematic in areas with narrow fringe spacing.
- Distortions caused by the combination of speckle and the subsequent homomorphic filtering manifested themselves as artefacts and gaps in the contours at the end of the

processing. This created the need for various automatic and semiautomatic editing strategies.

- The problem of non-linear modulation depth rendered inappropriate the use of global thresholding for fringe maxima segmentation. Application of this scheme produced residual artefacts such as erroneous bridges and gaps in the segments. Local thresholding, using 64 by 64 pixel regions proved to be much more effective for the segmentation in terms of reducing the number of artefacts.
- An obvious choice of a threshold for the segmentation was found to be problematic due to the intensity histogram shape of the nominally binary interference imagery. To solve the problem, the histogram centre of mass was chosen as the thresholding criterion.
- The contour image produced by the morphological thinning step contained many artefacts such as spur and bridge contours as well as gaps in contours. Automated editing procedures were successfully applied to remove many of the artefacts. These methods simply enforced the topological rules outlined in Chapter Four. For example, spur and bridge contours were eliminated by the simple fact that contour lines may not intersect.
- Some semiautomatic editing was required to compliment the automatic artefact removal. In general, areas with narrow fringe spacing required most of the semiautomatic editing.
- The success of the semiautomatic editing method of concentric contour generation was dependent upon the smoothness of the innermost contour and the shape

consistency of the concentric curves. In one image, several applications of this method were possible, while only one application was found to be appropriate in two other images.

- For two of the images processed, the efficiency of the new fringe maxima extraction algorithm was better than ninety per cent. Efficiency was quantified by the ratio of automatically extracted contour nodes to the total number of nodes. For the other two images, the efficiency metric "seemed" to indicate that the new method was not as successful. In this regard, the efficiency measure was somewhat misleading, as the number of nodes belonging to contours interpolated by semiautomatic means was not representative of the number of points manually delineated. Since only a maximum of three points per contour (for quadratic interpolation) were necessary, the new method was indeed more efficient than manual digitising than the efficiency measured suggested.
- An empirical study of the planimetric accuracy of the extracted contours was also conducted in order establish if there was consistency between manually digitised and automatically extracted fringe maxima contours. The testing yielded RMSE accuracy estimates ranging from ±1.1 to ±3.2 pixels. These figures were clearly dependent upon the nature of the fringe patterns, as the RMSE was lower for straight, narrow fringes and greater for curved and broad fringes.
- Application of the new method to a simulated image proved to be successful. However, there were noticeable artefacts at the borders due to the thinning. It has

been concluded, though, that this phenomenon is of little concern given that homomorphic filtering generally corrupts image border regions.

- It was shown that the final step of the new, two-dimensional phase unwrapping algorithm actually reduces to a one-dimensional problem. This is a direct result of the unique approach of the algorithm, the basis of which is a topological description of a phase map.

8.3 Recommendations for Future Development

In the author's view, the most significant areas requiring further development are tabulated below.

- While the speckle noise reduction via homomorphic filtering was successful, significant distortion was introduced in some locations, particularly where the fringe spacing was relatively narrow. An adaptive filtering scheme in which the filter properties vary according to fringe wavelength may hold promise for *a priori* minimisation of such distortions
- In order to minimise the amount of semiautomatic editing and, therefore, user interaction, a means of automatically identifying contour inconsistencies is required. A graph theoretical approach, in which topological relationships can be more rigorously implemented and enforced, may be suitable for both the identification and restitution of such artefacts.
- There may also be merit in investigating an iterative approach to contour extraction. At the outset of such a strategy, only the strongest, most distinct fringe maxima

contours would be extracted. At successive iterations, these curves could be used to assist in the extraction of weaker, less distinct fringe contours.

- A minor issue is that of the distortions introduced at the border regions of the fringe imagery by the filtering. Due to the assumed spatial periodicity of an image when performing a 2-D FFT, frequency domain filtering permits contributions of adjacent periods to the kernel response at the image borders. The solution to this problem may be as simple as making the image larger than necessary and ignoring the distorted border.
- Real off-axis holographic imagery should be acquired for testing the two-dimensional phase unwrapping algorithm. In spite of the success realised with the noisy simulated imagery, real data would be the ultimate test of the new algorithm. The author suspects that some of the contour inconsistencies encountered in the fringe imagery processing may crop up in a similar form in real phase maps. However, some of the presented solutions to these problems may also be transferable to phase map imagery.

REFERENCES

Andrä, P., U. Mieth and W. Osten (1991). "Strategies for Unwrapping Noisy Interferograms in Phase-Sampling Interferometry." In: Werner P. Jüptner (Ed.), Industrial Applications of Holographic and Speckle Measuring Techniques. SPIE Proceedings vol. 1508. 12-13 March, 1991. 50-60.

Bayly, D. A. (1991). "Machinery Alignment Monitoring with an Electronic Theodolite System." Ph.D. dissertation. The University of Calgary.

Beeck, M.-A. (1992). "Pulsed Holographic Vibration Analysis on High-Speed Rotating Objects: Fringe Formation, Recording Techniques, and Practical Applications." Optical Engineering. 31 (3): 553-561.

Bone, D. J. (1991). "Fourier Fringe Analysis: the Two-Dimensional Phase Unwrapping Problem." Applied Optics. 30 (25): 3627-3632.

Bone, D. J., H.-A. Bachor and R. J. Sandeman (1986). "Fringe-Pattern Analysis Using a 2-D Fourier Transform." Applied Optics. 25 (10): 1653-1660.

Brandt, G. B. (1979). "Holographic Interferometry." In: H. J. Caulfield (Ed.), Handbook of Optical Holography. Academic Press: New York, NY. 638 pages: 463-502.

Breuckmann, B. and W. Thieme (1985). "Computer-Aided Analysis of Holographic Interferograms Using the Phase-Shift Method." Applied Optics. 24 (14): 2145-2149.

Brigham, E. O. (1988). The Fast Fourier Transform and its Applications. Prentice-Hall, Inc.: Englewood Cliffs, NJ.

Bruce, A., D. Donoho and H.-Y. Gao (1996). "Wavelet Analysis." IEEE Spectrum. 33 (10): 26-35.

Burt, P. J. (1984). "The Pyramid as a Structure for Efficient Computation." In Multiresolution Image Processing and Analysis. A. Rosenfeld, (Ed.). Springer-Verlag: Berlin. 6-35.

Canny, J. (1986). "A Computational Approach to Edge Detection." IEEE Transactions on Pattern Analysis and Machine Intelligence. PAMI-8 (6): 679-698.

Castleman, K. R. (1996). Digital Image Processing. Prentice-Hall, Inc.: Englewood Cliffs, NJ.

Caulfield, H. J. (1979). "Speckle." In: H. J. Caulfield (Ed.), Handbook of Optical Holography. Academic Press: New York, NY. 638 pages: 367-371.

Caulfield, H. J. and S. Lu (1970). The Applications of Holography. John Wiley & Sons, Inc.: New York, NY.

Dainty, J. C. (1984). "Introduction." In: J. C. Dainty (Ed.), Laser Speckle and Related Phenomena. Second enlarged edition. Springer-Verlag: Berlin. 342 pages: 1-7.

Dudgeon, D. E. and R. M. Mersereau (1984). Multidimensional Digital Signal Processing. Prentice-Hall, Inc.: Englewood Cliffs, NJ.

Fincham, W. H. and M. H. Freeman (1974). Optics. Eighth edition. Butterworths & Co. (Publishers) Ltd.: London, UK.

Fraser, C. S. (1985). "Photogrammetric Measurement of Thermal Deformation of a Large Process Compressor." Photogrammetric Engineering and Remote Sensing. 51 (10): 1569-1575.

Fraser, C. S. and J. Shao (1998). "Scale-Space Methods for Image Feature Modelling in Vision Metrology." Photogrammetric Engineering and Remote Sensing. 64 (4): 323-328.

Gladden, J. W. and D. Leighty (1979). "Recording Media." In: H. J. Caulfield (Ed.), Handbook of Optical Holography. Academic Press: New York, NY. 638 pages: 277-298.

Goldstein, R. M., H. A. Zebker and C. L. Werner (1988). "Satellite Radar Interferometry: Two-Dimensional Phase Unwrapping." Radio Science. 23 (4): 713-720.

Gonzalez, R C. and R. E. Woods (1992). Digital Image Processing. Addison-Wesley Publishing Company: Reading, Ma.

Goodman, J. W. (1976). "Some Fundamental Properties of Speckle." Journal of the Optical Society of America. 66 (11): 1145-1150.

Goodman, J. W. (1984). In: J. C. Dainty (Ed.), Laser Speckle and Related Phenomena. Second enlarged edition. Springer-Verlag: Berlin. 342 pages: 9-75.

Greivenkamp, J. E. (1984). "Generalized Data Reduction for Heterodyne Interferometry." Optical Engineering. 23 (4): 350-352.

Greivenkamp, J. E. (1987). "Sub-Nyquist Interferometry." Applied Optics. 26 (24): 5245-5258.

Halliday, D. and R. Resnick (1988). Fundamentals of Physics. Third edition extended. John Wiley & Sons, Inc.: New York, NY.

Hamming, R. W. (1989). Digital Filters. Third edition. Prentice-Hall, Inc.: Englewood Cliffs, NJ.

Haralick, R. M. and L. G. Shapiro (1992). Computer and Robot Vision vol. 1. Addison-Wesley Publishing Company: Reading, MA.

Harris, F. J. (1987). "Multirate FIR Filters for Interpolating and Desampling." In: Douglas F. Elliott (Ed.), Handbook of Digital Signal Processing. Academic Press, Inc.: San Diego, CA. 999 pages: 173-287.

Hayt, W. H. Jr. (1989). Engineering Electromagnetics. Fifth edition. McGraw-Hill Publishing Company: New York, NY.

Hecht, J. (1992). Understanding Lasers: an Entry Level Guide. IEEE Press: New York, NY.

Hellwich, O. (1998). "SAR Phase Unwrapping Using Adaptive Recursive Smoothing." In: T. Schenk and A. Habib, (Eds.), Object Recognition and Scene Classification from Multispectral and Multisensor Pixels. ISPRS Proceedings vol. 32 part 3/1. 6-10 July, 1998. 492-500.

Hogg, R. V. and A. T. Craig (1959). Introduction to Mathematical Statistics. The Macmillan Company: New York, NY.

Holst, G. C. (1996). CCD Arrays, Cameras, and Displays. JCD Publishing: Winter Park, FL, and SPIE Optical Engineering Press: Bellingham, WA.

Huntley, J. M. (1989). "Noise-Immune Phase Unwrapping Algorithm." Applied Optics. 28 (15): 3268-3270.

Huntley, J. M. and H. Saldner (1993). "Temporal Phase-Unwrapping Algorithm for Automated Interferogram Analysis." Applied Optics. 32 (17): 3047-3052.

Itoh, K. (1982). "Analysis of the Phase Unwrapping Algorithm." Applied Optics. 21 (4): 2470.

Judge, T. R. and P. J. Bryanston-Cross (1994). "A Review of Phase Unwrapping Techniques in Fringe Analysis." Optics and Lasers in Engineering. 21: 199-239.

Judge, T. R., C. Quan and P. J. Bryanston-Cross (1992). "Holographic Deformation Measurements by Fourier Transform Technique with Automatic Phase Unwrapping." Optical Engineering. 31 (3): 533-543.

Kaplan, W. (1984). Advanced Calculus. Third edition. Addison-Wesley Publishing Company: Redwood City, CA.

Kölbl, O. (1994). "Survey of High-Quality Photographic Scanners." In: Proceedings: Mapping and Remote Sensing Tools for the 21st Century. 26-29 August, 1994, Washington, DC. 7-14.

Kopeika, N. S. (1998). A System Engineering Approach to Imaging. SPIE Optical Engineering Press: Bellingham, WA.

Kreis, T. (1986). "Digital Holographic Interference-Phase Measurement Using the Fourier Transform Method." Journal of the Optical Society of America. 3 (6): 847-855.

Kujawinska, M. (1987). "Use of Phase-Stepping Automatic Fringe Analysis in Moiré Interferometry." Applied Optics. 26 (22): 4712-4714.

Lam, K. (1990). "An Object Based Image Matching Procedure for Photogrammetric Applications." M.Sc. thesis. The University of Calgary

Lathi, B. P. (1992). Linear Systems and Signals. Berkeley-Cambridge Press: Carmichael, CA.

Lee, W.-H. (1972). "Effect of Film-Grain Noise on the Performance of Holographic Memory." Journal of the Optical Society of America. 62 (6): 797-801.

Li, D. and J. Shao (1994). "The Wavelet and its Application in Image Edge Detection." ISPRS Journal of Photogrammetry and Remote Sensing. 49 (3): 4-11.

Lim, J. S. and H. Nawab (1981). "Techniques for Speckle Noise Removal." Optical Engineering. 20 (3): 472-480.

Macy, W. W. (1983). "Two-Dimensional Fringe-Pattern Analysis." Applied Optics. 22 (23): 3898-3901.

Malhotra, R. C. (1970). "Holography as Viewed by a Photogrammetrist." Photogrammetric Engineering. 36 (2): 152-162.

McKechnie, T. S. (1984). "Speckle Reduction." In: J. C. Dainty (Ed.), Laser Speckle and Related Phenomena. Second enlarged edition. Springer-Verlag: Berlin. 342 pages: 123-170.

McKinney, R. G., P. Z. Aldestein, M. G. Anderson, E. C. Doerner, A. E. Fields, N. L. Fritz, J. Gravelle, J. F. Hamilton, G. L. Robinson, J. T. Smith Jr., M. R. Specht, E. G. Tibbils, W. F. Voglesong and R. A. Welch (1980). "Photographic Materials

and Processing." In: C. C. Slama (Ed.), Manual of Photogrammetry. Fourth edition. American Society of Photogrammetry: Falls Church, VA. 1056 pages: 305-366.

Mikhail, E. M. and G. H. Glaser (1971). "Mensuration Aspects of Holograms." Photogrammetric Engineering. 37 (3): 267-276.

Moore, A. J. and J. R. Tyrer (1995). "Phase-Stepped ESPI and Moiré Interferometry for Measuring Stress-Intensity Factor and *J* Integral." Experimental Mechanics. 35 (4): 306-314.

Nevatia, R. and K. R. Babu (1980). "Linear Feature Extraction and Description." Computer Graphics and Image Processing. 13: 257-269.

Nugent, K. A. (1985). "Interferogram Analysis Using an Accurate Fully Automatic Algorithm." Applied Optics. 24 (18): 3101-3105.

Oppenheim, A. V. and R. W. Schafer (1989). Discrete-Time Signal Processing. Prentice-Hall, Inc.: Englewood Cliffs, NJ.

Papoulis, A. (1962). The Fourier Integral and its Applications. McGraw-Hill Publishing Company: New York, NY.

Papoulis, A. (1991). Probability, Random Variables and Stochastic Processes. Third edition. McGraw-Hill Publishing Company: New York, NY.

Pascal, J. –C., X. Carniel, V. Chalvidan and P. Smigielski (1996). "Determination of Phase and Magnitude of Vibration for Energy Flow Measurements in a Plate Using Holographic Interferometry." Optics and Lasers in Engineering. 25: 343-360.

Pedrotti, F. L. and L. S. Pedrotti (1993). Introduction to Optics. Second edition. Prentice-Hall Inc.: Englewood Cliffs, NJ.

Poon, C. Y., M. Kujawinska and C. Ruiz (1993a). "Spatial Carrier Phase Shifting Method of Fringe Analysis for Moiré Interferometry." Journal of Strain Analysis for Engineering Design. 28 (2): 79-88.

Poon, C. Y., M. Kujawinska and C. Ruiz (1993b). "Automated Fringe Pattern Analysis for Moiré Interferometry." Experimental Mechanics. 33 (3): 234-241.

Pouet, B. and S. Krishnaswamy (1996). "Dynamic Holographic Interferometry by Photorefractive Crystals for Quantitative Deformation Measurements." Applied Optics. 35 (5): 787-794.

Pratt, W. K. (1976). Digital Image Processing. John Wiley & Sons, Inc.: New York, NY.

Pryputniewicz, R. J. (1989). "Hologrammetry: Systems and Applications." In: H. M. Karara (Ed.), Non-Topographic Photogrammetry. Second edition. American Society for Photogrammetry and Remote Sensing: Falls Church, VA. 445 pages: 203-230.

Quan, C., H. M. Shang and P. J. Bryanston-Cross (1996). "Application of the Holographic Carrier Fringe and FFT Technique for Deformation Measurement." Optics & Laser Technology. 28 (1): 7-13.

Ray, S. F. (1994). Applied Photographic Optics. Second edition. Focal Press: Oxford, UK.

Reid, G. T. (1986). "Automatic Fringe Pattern Analysis: A Review." Optics and Lasers in Engineering. 7: 37-68.

Rioul, O. and M. Vetterli (1991). "Wavelets and Signal Processing." IEEE Signal Processing Magazine. 8 (4): 14-38.

Robinson, D. W. and G. T. Reid (eds.) (1993). Interferogram Analysis: Digital Fringe Pattern Measurement Techniques. Institute of Physics Publishing: Bristol, UK.

Schenk, T. (1995). "Scale Space Techniques for Mobile Mapping Imagery." In: Proceedings 1995 Mobile Mapping Symposium. 24-26 May, 1995. 28-34.

Schmidt, T. and J. W. Webster (1996). "Portable Holographic Interferometry Camera Detects Deterioration in Ageing Aircraft." Sound and Vibration. June: 2-6.

Schwinder, J., R. Burow. K.-E. Elssner, J. Grzanna, R. Spolaczyk and K. Merkel (1983). "Digital Wave-Front Measuring Interferometry: Some Systematic Error Sources." Applied Optics. 22 (21): 3421-3432.

Shao, J. and W. Förstner (1994). "Gabor Wavelets for Texture Edge Extraction." International Archives of Photogrammetry. Vol. 30 no. 3. 745-752.

Snell, J. F. (1978). "Radiometry and Photometry." In: Walter G. Driscoll (Ed.), Handbook of Optics. McGraw-Hill Book Company: New York, NY. 1-30.

Stefanidis, A. and T. Schenk (1992). "On the Application of Scale Space Techniques in Digital Photogrammetry." International Archives of Photogrammetry. Vol. 29 part B3 comm. 3. 586-590.

Strang, G. (1989). "Wavelets and Dilation Equations: A Brief Introduction." SIAM Review. 31 (4): 614-627.

Strang, G. (1993). "Wavelet Transforms Versus Fourier Transforms." Bulletin of the American Mathematical Society. 28 (2): 288-305.

Takeda, M., H. Ina and S. Kobayashi (1982). "Fourier-Transform Method of Fringe-Pattern Analysis for Computer-Based Topography and Interferometry." Journal of the Optical Society of America. 72 (1): 156-160.

Thompson, B. J. (1979). "Interference and Diffraction." In: H. J. Caulfield (Ed.), Handbook of Optical Holography. Academic Press: New York, NY. 638 pages: 29-41.

Tribolet, J. M. (1977). "A New Phase Unwrapping Algorithm." IEEE Transactions on Acoustics, Speech and Signal Processing. ASSP-25 (2): 170-177.

Vest, C. M. (1979). Holographic Interferometry. John Wiley & Sons, Inc.: New York, NY.

Vilkomerson, D. (1971). "Statistical Distribution of Irradiance in the Creation of a Hologram." Journal of the Optical Society of America. 61 (7): 929-933.

Williams, J. R. and K. Amaratunga (1994). "Introduction to Wavelets in Engineering." International Journal for Numerical Methods in Engineering. 37 (14): 2365-2388.

Yatagai, T., S. Nakadate, M. Idesawa and H. Saito (1982). "Automatic Fringe Analysis Using Digital Image Processing Techniques." Optical Engineering. 21 (3): 432-435.

Yu, F. T. S. and A. Tai (1979). "Image Evaluation." In: H. J. Caulfield (Ed.), Handbook of Optical Holography. Academic Press: New York, NY. 638 pages: 51-68.

APPENDIX A

RELEVANT IMAGE TRANSFORMS AND THEOREMS

Transforms are versatile tools used extensively in the broad field of image processing to provide a different basis for analysing functions. The Fourier transform, for example, gives an equivalent sum-of-sinusoids representation of a function. The benefit of this representation is that significant periodic components of the function are more readily apparent in the frequency domain than in the spatial domain. Both the continuous and discrete versions of the two-dimensional Fourier transform along with some relevant theorems are presented in this appendix.

The following discussions represent a focused treatment of image transforms relevant to this research. The process of phase recovery, which constitutes important foundation material for this research, draws upon some of the continuous Fourier transform theorems. In addition, practical phase recovery from digital imagery necessitates use of the discrete Fourier transform. It will be shown that the practical implementation of digital filters can often be performed more efficiently in the frequency domain. This is particularly relevant since frequency domain homomorphic filtering is an integral part of the fringe maxima extraction algorithm developed herein. References drawn upon for this appendix include Brigham (1988), Castleman (1996), Dudgeon and

Mersereau (1984), Gonzalez and Woods (1992), Oppenheim and Schafer (1989) and Papoulis (1962).

An increasingly popular and versatile alternative to the Fourier transform is the wavelet transform. The difference between the two lies in their basis functions. While the basis functions of the Fourier transform are infinite-duration, complex sinusoids, the wavelet transform employs short-duration oscillatory waveforms. The advantage of the latter transform lies in the analysis and detection of transient components of a signal. This is particularly relevant for edge detection, used here for phase recovery. Edge detection is essentially a search for (transient) discontinuities in image intensity. The wavelet transform and the closely related subject of multi-resolution image representation and scale space are treated at the end of this appendix. Both these subjects are employed in this thesis for phase map discontinuity detection and extraction.

A.1 The Two-dimensional Continuous Fourier Transform

The two-dimensional continuous Fourier transform (2-D CFT) is a linear, integral transform that maps a function from the spatial domain onto the spatial frequency domain. The transform is given by

$$F(\omega,\nu)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\, e^{-j(\omega x+\nu y)}dxdy \ , \qquad (A.1)$$

where

$f(x,y)$ is the spatial domain function,

$F(\omega,\nu)$ is the spatial frequency domain function, or spectrum.

The units of the frequency variables, ω and ν, are radians per linear unit. Both the spatial domain function and the spatial frequency domain function, if it exists, are complex-valued in general. The inverse 2-D CFT is simply a mapping of a frequency domain function back to the spatial domain, and is given by

$$f(x,y)=\frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}F(\omega,\nu)\,e^{j(\omega x+\nu y)}d\omega d\nu \ . \tag{A.2}$$

Both of these integrals are exploited for phase map recovery in Chapter Four. The relation between a function and its Fourier transform is often denoted symbolically as

$$f(x,y)\Leftrightarrow F(\omega,\nu) \ . \tag{A.3}$$

A.2 Two-dimensional Continuous Fourier Transform Theorems

An important theorem in signal modulation and demodulation is the frequency-shifting or modulation theorem that gives the transform pair for a spatial domain function multiplied by a complex sinusoid. The result in the Fourier domain is translation of the spectrum:

$$e^{j(\omega_o x+\nu_o y)}f(x,y)\Leftrightarrow F(\omega-\omega_o,\nu-\nu_o) \ . \tag{A.4}$$

Thus, a frequency shift causes a phase shift in the spatial domain function. The spatial-differentiation theorem is used to evaluate the spectrum of a differentiated function:

$$\left(\frac{\partial}{\partial x}\right)^m\left(\frac{\partial}{\partial y}\right)^n f(x,y)\Leftrightarrow (j\omega)^m(j\nu)^n F(\omega,\nu) \ . \tag{A.5}$$

If a function is differentiated m times in the spatial domain with respect to x, its spectrum is multiplied by $j\omega$ raised to the power of m. If a function and its Fourier transform are

related by Equation A.3, then the transform pair corresponding to the complex conjugate of that function is given by

$$f^*(x,y) \Leftrightarrow F^*(-\omega,-\nu) \; . \tag{A.6}$$

Chapter Four makes use of the modulation and complex conjugate theorems, while reference to the differentiation theorem is made in Chapter Six.

A.3 The Two-dimensional Discrete-Space Fourier Transform

In order to facilitate digital manipulation, two-dimensional functions must be discretely sampled. The most common uniform sampling geometry is a rectangular lattice, though other patterns such as hexagonal sampling are used in some instances. Although the two-dimensional sampling process is not detailed here, the frequency domain representation of a discrete function is presented. For discussion on sampling, see Dudgeon and Mersereau (1984) or Oppenheim and Schafer (1989).

A two-dimensional discrete function may be represented as $f(m,n)$, where m and n are integer-valued indices. The discrete-space Fourier transform of $f(m,n)$ is given by

$$F(e^{j\omega},e^{j\nu}) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} f(m,n)\, e^{-j(\omega m+\nu n)} \; . \tag{A.7}$$

A distinction must be made between the frequency variables ω and ν of this equation and those of Equation A.1. The units of the latter are radians per linear unit, while for the discrete-space transform, the units are radians. Moreover, the complex exponential notation for the independent variables of Equation A.7 is used because the spectrum of a discrete function is periodic with period 2π. It can be shown that, if $f(m,n)$ is a discrete

version of $f(x,y)$, Equation A.7 is a periodic replication of the continuous spectrum $F(\omega,\nu)$.

A.4 The Two-dimensional Discrete Fourier Transform

In addition to a discrete version of a spatial signal, it is desirable to also have a sampled version of the corresponding spectrum. It can be shown that such a representation is obtained if periodicity of the discrete space function is assumed. Under this assumption, the two-dimensional discrete Fourier transform (2-D DFT) of $f(m,n)$ is given by

$$F(u,v)=\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}f(m,n)e^{-j2\pi\left(\frac{mu}{M}+\frac{nv}{N}\right)} \quad . \tag{A.8}$$

There are two interpretations of this transform. First, it is a discretely sampled version of the spectrum of $f(m,n)$, given by Equation A.7. Or, it is the array of two-dimensional discrete Fourier series coefficients for the rectangularly periodic function $\tilde{f}(m,n)$, with spatial periods M and N, of which $f(m,n)$ constitutes one period. The relation between these two functions is

$$\tilde{f}(m,n)=\sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty}f(m+pM,n+qN) \; . \tag{A.9}$$

The inverse 2-D DFT is given by

$$f(m,n)=\frac{1}{MN}\sum_{u=0}^{M-1}\sum_{v=0}^{N-1}F(u,v)e^{j2\pi\left(\frac{mu}{M}+\frac{nv}{N}\right)} \quad . \tag{A.10}$$

Because the two-dimensional Fourier kernel is separable, Equation A.8 can be decomposed into

$$F(u,v) = \sum_{m=0}^{M-1} e^{-j2\pi\frac{mu}{M}} \sum_{n=0}^{N-1} f(m,n) e^{-j2\pi\frac{nv}{N}} \quad . \tag{A.11}$$

Under this decomposition, each column of $f(m,n)$ is transformed, then each row is transformed, or vice-versa.

A.5 The Fast Fourier Transform

Equations A.8 and A.10 facilitate digital computation of the forward and inverse Fourier transforms of a two-dimensional function. However, even if the simple decomposition of Equation A.11 is utilised, application of these formulae entails a great deal of numerical computation. The fast Fourier transform (FFT) is a powerful factorisation algorithm that significantly reduces the number of computations required. To illustrate the algorithm, which was used extensively in this research, the factorisation will be performed on the one-dimensional DFT, and then the generalisation to the 2-D DFT will be made.

The form of the one-dimensional DFT (1-D DFT) is essentially identical to that of Equation A.8, just with one less dimension:

$$F(u) = \sum_{n=0}^{N-1} f(n) e^{-j\frac{2\pi mu}{N}} \quad . \tag{A.12}$$

One particular FFT algorithm, the radix-2 decimation-in-time algorithm, hinges on the requirement that the number of samples in $f(n)$, N, is an integer power of 2, i.e.,

$$N = 2^x \quad . \tag{A.13}$$

Using this relation, N is set equal to 2M for the first decomposition of Equation A.12. The first step of the decomposition is accomplished by expanding the series and then grouping together the even and odd index terms

$$F(u) = \sum_{n=0}^{2M-1} f(n) e^{-j\frac{2\pi nu}{2M}}$$

$$= \sum_{n=0}^{M-1} f(2n) e^{-j\frac{2\pi nu}{M}} + e^{-j\frac{\pi u}{M}} \sum_{n=0}^{M-1} f(2n-1) e^{-j\frac{2\pi nu}{M}} \tag{A.14}$$

The result is two transforms, each of length M rather than one transform of length 2M. The even terms are given by the first summation term, while the odd terms are contained in the second term. The decomposition is then performed on each of the two series of Equation A.14, resulting in four transforms, each of length $M/2$. This process is repeated until only $N/2$ transforms of length 2 need be performed. By using this algorithm, the number of complex computations is reduced from N^2 to $N \log_2 N$, which represents considerable savings for large signals (Gonzalez and Woods, 1993). The extension of the FFT to two-dimensions is simply a matter of using the 1-D FFT algorithm on the decomposed form of the 2-D DFT given by Equation A.11.

A.6 Two-dimensional Discrete Fourier Transform Theorems

The only 2-D DFT theorem exploited during the course of this research was the convolution theorem. Simply stated, the convolution of two functions in the spatial

domain is equivalent to their multiplication in the frequency domain. The benefit of this theorem is reaped when using a digital filter, $h(m,n)$, with large support. Substantial computational savings are realised by utilising the FFT algorithm performing the filtering in the frequency domain. The theorem is given by

$$f(m,n)**h(m,n) \Leftrightarrow F(u,v)H(u,v) \; . \tag{A.15}$$

Use is made of this theorem in conjunction with the 2-D FFT for homomorphic filtering in Chapters Six and Seven.

A.7 Multi-resolution Representation and the Wavelet Transform

As mentioned at the outset of this appendix, the Fourier transform provides an alternate representation of a function on the basis of complex sinusoids. The drawback of using this transform is the basis functions are infinite in duration, and are not well suited for analysing transient phenomena such as edges in imagery (Rioul and Vetterli, 1991; Bruce *et al.*, 1996). While the basis functions of the wavelet transform (WT) are also oscillatory, they are of shorter duration. Moreover, the WT provides a representation of a function utilising several scaled versions of the basis functions, the wavelets. The benefit of this approach is that both long-period phenomena (global trends in an image) as well as short-period trends (image details) are captured by the WT. This characteristic is exploited for discontinuity detection in Chapter Six.

The idea of a multiple-scale or multiple-resolution representation of an image is closely related to the function of the wavelet transform. A scale space representation provides views of an image at different scales, and is formed by filtering the image with

operators of increasing support (scale). At low scales, the image details are captured with narrow filters, while the broad filters at larger scales contain the global trends of the signal. While scale space theory allows for a continuous scaling parameter, it is usually discretized into integer powers of two for image analysis. The scales of any two successive levels of a scale space representation thus differ by a power of two. In terms of dimensions, a multi-scale image representation can be implemented in one of two ways: pyramidal and parallel (Schenk, 1995). In the parallel implementation, the image dimensions at each scale are the same as those of the original image. In the pyramidal representation, decimation is performed after the filtering, thereby reducing the dimensions by a factor of two in each direction (x and y). A Gaussian pyramid is a multi-resolution representation of an image comprised of low pass filtered and decimated versions of the original image (Burt, 1984). The name stems from the use of a Gaussian kernel as an approximation to the impulse response of an ideal low pass filter.

The two-dimensional continuous wavelet (2-D CWT) transform representation of a function is given by the following integral (Li and Shao, 1993):

$$Wf(a,\tau_x,\tau_y)=\frac{1}{\sqrt{|a|}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)\psi^{*}\left(\frac{x-\tau_x}{a},\frac{y-\tau_y}{a}\right)dxdy\,, \qquad (A.16)$$

where

$Wf(a,\tau_x,\tau_y)$ is the wavelet representation of the function $f(x,y)$,

a is the scale or dilation parameter,

τ_x, τ_y are the x- and y- translation parameters, respectively,

$\psi(x,y)$ is the basic wavelet, from which all wavelets are derived.

$\psi\left(\frac{x-\tau_x}{a},\frac{y-\tau_y}{a}\right)$ is the wavelet, a translated and dilated version of the basic wavelet.

Note that the scale and translation parameters are real-valued. Equation A.16 can be written in short form as the convolution of the function and the basic wavelet (Li and Shao, 1993)

$$Wf(a,\tau_x,\tau_y)=f(\tau_x,\tau_y)**\psi(-\tau_x,-\tau_y)\ . \tag{A.17}$$

The scale parameter is often expressed in terms of powers of two

$$a=2^j\ , \tag{A.18}$$

where j is real.

The choice of a basis function, the basic wavelet, is very much application-dependent. The choice is somewhat restricted, though, as a wavelet must satisfy the zero mean condition

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\psi(x,y)dxdy=0\ . \tag{A.19}$$

A good wavelet has a compact region of support (finite extent) in the spatial domain. Another property of a good basic wavelet is that it is orthogonal to its translates and dilates (translated and scaled wavelets).

The first derivative of a Gaussian function is a wavelet based upon the criteria of Equation A.19 (Li and Shao, 1993). Although this function lacks compact support, for practical purposes it can be considered to be of finite duration. This wavelet does not exhibit the orthogonality property, though. Further justification for this choice of basic wavelet, based upon its suitability for discontinuity detection, is given in Chapter Six.

The one-dimensional version of the first derivative of a Gaussian function is given by

$$\psi(x)=\frac{d}{dx}\left\{\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2}\right\},$$

$$=\frac{-x}{\sqrt{2\pi}\sigma^3}e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2} \qquad \text{(A.20)}$$

where the standard deviation, σ, is treated as the scale parameter:

$$\sigma=2^j\ . \qquad \text{(A.21)}$$

A plot of this basic wavelet and its dilates for various values of the scale parameter is given in Figure A.1. Note that as the scale parameter, increases, the wavelet width increases and the maximum amplitude decreases. A major benefit to using the wavelet transform is the fact that the filter width is matched to the scale. The peaks occur at $x=\pm\sigma$ and the only zero crossing is located at $x=0$. Also note that the curves exhibit odd symmetry and asymptotically approach zero as $x\to\pm\infty$.

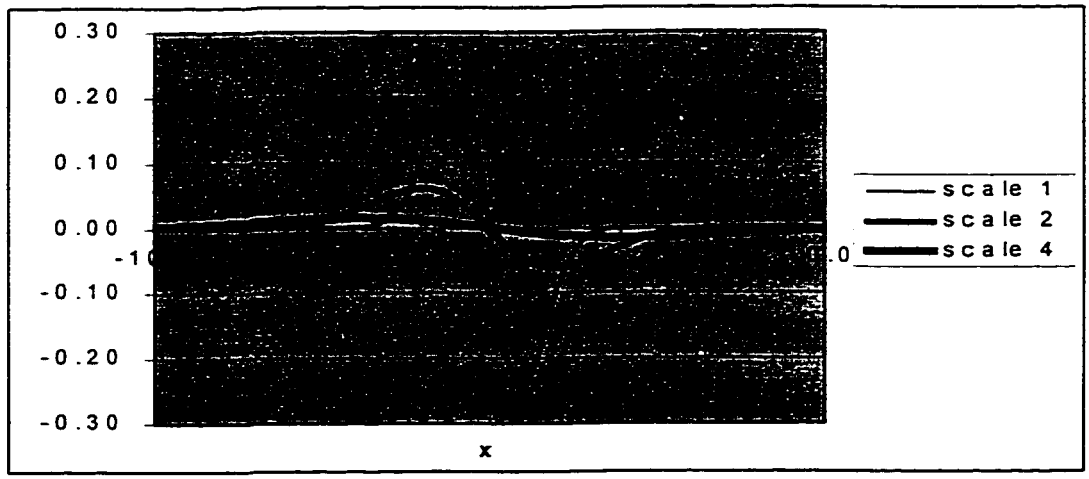


Figure A.1. One-dimensional First Derivative of a Gaussian Wavelets.

In two dimensions, there are two basic wavelets corresponding to the derivatives of the circularly-symmetric Gaussian function taken with respect to the independent variables x and y

$$\psi^x(x,y)=\frac{\partial}{\partial x}\left\{\frac{1}{2\pi\sigma^2}e^{-\left(\frac{x^2+y^2}{2\sigma^2}\right)}\right\},$$

$$=\frac{-x}{2\pi\sigma^4}e^{-\left(\frac{x^2+y^2}{2\sigma^2}\right)} \qquad \text{(A.22)}$$

and

$$\psi^y(x,y)=\frac{\partial}{\partial y}\left\{\frac{1}{2\pi\sigma^2}e^{-\left(\frac{x^2+y^2}{2\sigma^2}\right)}\right\}.$$

$$=\frac{-y}{2\pi\sigma^4}e^{-\left(\frac{x^2+y^2}{2\sigma^2}\right)} \qquad \text{(A.23)}$$

These equations represent the wavelets used for edge detection in wrapped phase maps.