Temporal phase unwrapping: development and application of real-time systems for surface profile and surface displacement measurement

This item was submitted to Loughborough University's Institutional Repository by the/an author.

Additional Information:

- A Doctoral Thesis. Submitted in partial fulfillment of the requirements for the award of Doctor of Philosophy of Loughborough University.

Metadata Record: https://dspace.lboro.ac.uk/2134/6824

Publisher: © Charles Russell Coggrave

Please cite the published version.
This item is held in Loughborough University’s Institutional Repository (https://dspace.lboro.ac.uk/) and was harvested from the British Library’s EThOS service (http://www.ethos.bl.uk/). It is made available under the following Creative Commons Licence conditions.

For the full text of this licence, please go to:
http://creativecommons.org/licenses/by-nc-nd/2.5/
Temporal Phase Unwrapping: Development and Application of Real-Time Systems for Surface Profile and Surface Displacement Measurement

by

Charles Russell Coggrave

A Doctoral Thesis
Submitted in partial fulfilment of the requirements for the award of Doctor Of Philosophy of Loughborough University

August, 2001

© by Charles Russell Coggrave, 2001
Acknowledgements

This work has been supported in part by EPSRC grant GR/K67274 entitled “Real-time low-noise speckle interferometry system for displacement strain and surface profile measurement”. I would also like to acknowledge financial support from Brown & Sharpe to study the optimisation of profilometry systems based on spatial light modulators. I am grateful for guidance received from Dr Philip Aldred regarding the requirements of commercial co-ordinate measurement systems that was invaluable during this study. Design of the support structure for the shape measurement apparatus was achieved in collaboration with Brown & Sharpe; in particular the contributions of Steve Martin and Daniel Cosandey are acknowledged.

I would also like to acknowledge the contribution of Dr. Phillip Salter of BAe Systems for influential discussions on speckle metrology applications in the international aerospace industry. David Panni kindly loaned carbon fibre samples that were used to validate the real-time out-of-plane digital speckle pattern interferometry system. Construction and use of interferometer arrangements within the Structural Integrity Research group tends to be of a collaborative nature, due in part to the high cost of optical components, and I would therefore like to thank all members of the team for their assistance. The contribution of Professor G.H. Kaufmann to the out-of-plane interferometer described in chapter 7 was particularly helpful.

Finally, I would like to thank my supervisor Professor J.M. Huntley for his friendship and enthusiastic academic support over the past five years.
Abstract

Industrial adoption of whole-field optical metrology has been hindered by the interpretation of the resulting wrapped phase data using traditional spatial unwrapping algorithms. These algorithms typically require long computation times and are often unable to provide unique solutions. An alternative approach described by Huntley and Saldner\(^1\) simplifies the data analysis by using a time-series of wrapped phase maps to unwrap the data temporally, and has the advantage that errors due to noise and specimen boundaries do not propagate spatially. However, viability of the algorithm is restricted by the need to store and subsequently analyse very large datasets. This thesis describes the development and application of an instrument that implements the temporal phase unwrapping algorithm (TPUA) in real-time to overcome these problems and allows whole-field optical techniques to become more intuitive by enabling quasi-live results to be displayed.

Tradeoffs across the multiple domains of algorithm, hardware and software are discussed, including decomposition of the algorithm onto particular architectures and implementation using a commercial pipeline image processing system.

In the first application, the surface profile of discontinuous objects is measured using a Digital Mirror Device spatial light modulator (SLM) to project an optimised sequence of sinusoidal white light fringes onto the object surface at 60 frames s\(^{-1}\). Less than 0.5 s is required to measure and display approximately 250,000 co-ordinates with a precision of better than one part in 5,000 of the field of view.

Issues affecting the performance of white light projected fringe profilometers implemented using SLMs are investigated. Defocusing of the projector is shown to be a critical limiting factor, with a precision of better than one part in 20,000 of the field of view being achieved when optimised.

A speckle interferometer is used in the second application to measure object displacement. Quasi-live unwrapped speckle interferograms are displayed at 15 frames s⁻¹ using either a piezoelectric transducer-mounted prism or a Pockels cell as the phase-stepping device. The reference speckle interferogram is automatically updated at regular intervals allowing arbitrarily large deformations to be measured. The signal-to-noise ratio of the calculated displacement fields can be improved by performing real-time temporal least-squares fitting.
# Table of contents

1 Whole-field optical metrology........................................................................................................... 2

1.1 Introduction........................................................................................................................................ 2

1.2 Surface displacement measurement................................................................................................. 3

1.2.1 Smooth wavefront interferometry ................................................................................................. 3

1.2.2 Holographic interferometry .......................................................................................................... 5

1.2.3 Speckle interferometry..................................................................................................................... 7

1.2.4 Out-of-plane speckle interferometry............................................................................................... 9

1.2.5 In-plane speckle interferometry .................................................................................................... 11

1.2.6 Speckle shearing interferometry .................................................................................................. 12

1.2.7 Difference techniques.................................................................................................................... 14

1.3 Surface profile measurement.............................................................................................................. 14

1.3.1 Introduction.................................................................................................................................... 14

1.3.2 Computer vision techniques.......................................................................................................... 16

1.3.3 Photogrammetry............................................................................................................................ 17

1.3.4 Contouring speckle interferometry ............................................................................................... 19

1.3.5 Low coherence interferometry...................................................................................................... 21

1.3.6 Fourier transform speckle profilometry......................................................................................... 22

1.3.7 Spectral interference microscope.................................................................................................. 24

1.3.8 Coaxial coimage plane projection and observation profilometry................................................ 25

1.3.9 Binary coding of projected structured light ................................................................................. 25

1.3.10 Spatial frequency scanning of projected structured light............................................................ 27

1.3.11 Multiplexed spatial frequency of projected structured light....................................................... 30

1.4 Summary........................................................................................................................................ 30

1.5 Figures........................................................................................................................................... 33

2 Quantitative interferogram analysis.................................................................................................. 41

2.1 Introduction....................................................................................................................................... 41

2.2 Temporal phase shifting.................................................................................................................... 45

2.2.1 Four-frame algorithm..................................................................................................................... 47

2.2.2 Five-frame algorithm...................................................................................................................... 48

2.2.3 Carré technique.............................................................................................................................. 48
Scope of thesis

Introduction

There is a recognised industrial requirement for experimental techniques that enable the basic physical and mechanical properties of materials to be measured. The field of non-destructive testing has become increasingly important in many high-technology industrial areas (e.g. automotive, aerospace, and semiconductor fabrication), and may form part of a quality control procedure, which in the case of safety-critical components can involve rigorous tests on each sample. The same techniques are often applied at the design stage of complex components where experimental data is required to validate theoretical performance models. Conventional measurement methods such as resistance strain gauging and stylus profilometry have several drawbacks: (a) they typically involve contact with the test specimen and may therefore modify the physical quantity of interest; (b) they provide only local information and require many individual measurements, either in sequence or in parallel, to collect whole-field data; and (c) are labour intensive to apply.

A wide variety of whole-field optical techniques for non-contacting measurement of mechanical properties have been developed over the last 30 years, however, they have seen only limited industrial adoption during this period. Widespread application of these techniques has been held back by the need to convert the measured intensity data (so-called interferograms) into a form that describes the physical quantity of interest, and that can be understood and applied by engineers who may not be skilled in optics. The main breakthrough to a more widespread range of applications came in 1985 with the introduction of the phase-stepping method to speckle interferogram analysis[1,2]. This enables two-dimensional phase maps corresponding to the quantity of interest to be extracted directly from interferograms measured by a video camera interfaced to a computer. The resulting phase values lie in the principle range \([-\pi, \pi]\) and must be “unwrapped” before the physical parameter of interest can be calculated. The conventional approach to unwrapping involves adding integral multiples of \(2\pi\) to all pixels such that the phase difference between neighbours falls in the range \([-\pi, \pi]\). This simple spatial comparison procedure can give rise to
significant errors when analysing noisy data and several extensions to the basic method have been proposed. Unfortunately, most of these methods are computationally intensive and often require expert knowledge to define the boundary conditions.

This thesis is concerned with automated quantitative analysis of data measured using whole-field optical methods, and investigates the feasibility of presenting real-time results that can be interpreted by the non-specialist. A novel inspection system is developed that overcomes some of the issues limiting industrial adoption of optical methods as follows: (a) phase-stepped interferograms are acquired at normal video rates to reduce the effects of low-frequency vibration and environmental disturbances typically found at industrial sites; (b) data are processed on-line so that intermediate results are hidden from the end-user, thus simplifying the interpretation of results; and (c) results are displayed on a computer monitor in real-time, thereby providing immediate feedback and enabling the measurement procedure to become more intuitive.

The measurement procedure is based on an alternative approach to phase unwrapping[3], and assembles a sequence of two-dimensional phase maps to form a three-dimensional wrapped phase distribution. Data analysis is intrinsically computationally efficient, and involves one-dimensional comparison of phase values along the time axis, rather than two-dimensional spatial comparison as described above. In so-called temporal phase unwrapping, each pixel is unwrapped independently of all others so that physical discontinuities and regions of poor signal to noise ratio do not propagate spatially to adversely influence good data points. In general, specification of boundary conditions is not required when using the temporal phase unwrapping algorithm, thus removing the requirement for expert user interaction.

Application of the instrument to: (a) measurement of surface profile by projection of sinusoidal fringes; and (b) measurement of surface displacement using digital speckle pattern interferometry techniques; is investigated.
Original contribution
The work presented in this thesis is original, except where otherwise stated, and includes the following fundamental issues:

(i) Investigation of the temporal phase unwrapping algorithm mapping to several data processing architectures, and a study of the data throughput rate that can be achieved using an optimised algorithm implementation for the von Neumann architecture. The investigation identifies that commercial pipeline image processors offering flexible hardware architectures and deterministic data throughput are well suited for implementation of the temporal phase unwrapping algorithm.

(ii) Development of a profilometry system based on projection of white-light sinusoidal fringes, that computes co-ordinate information in real-time using an extended form of the temporal phase unwrapping algorithm implemented on a pipeline image processing architecture.

(iii) Methodical experimental investigation of systematic error sources in profilometers based on the projection of white-light sinusoidal fringes generated using a spatial light modulator.

(iv) Development of a digital speckle pattern interferometry system that is able to compute and display two-dimensional unwrapped phase maps in real-time from speckle interferograms acquired at normal video rates. The system implements an extended temporal phase unwrapping algorithm that has been optimised for speckle data based on a pipeline image processing architecture, and has been demonstrated experimentally using out-of-plane interferometer arrangements.

Items (ii) and (iii) are the subject of papers appearing in peer-reviewed journals[4,5], and the author has been invited to submit item (iv) for publication based on the content of a recent conference presentation. A complete list of journal publications and papers in conference proceedings by the author is included in appendix A12.

Organisation of thesis
Chapter 1 provides an overview of whole-field optical techniques commonly used for measuring surface displacement and surface profile. The chapter aims to demonstrate
the breadth of research that is being applied in this field, and introduces the fundamental principles for each method.

Chapter 2 presents a review of automated methods for quantitative analysis of two-dimensional interferograms typically recorded using optical measurement techniques. Phase-stepping methods used to estimate the unknown phase corresponding to the physical quantity of interest are introduced, and unwrapping algorithms that remove the ambiguity in the phase signal are described.

Chapter 3 considers the tradeoffs involved in moving the temporal phase unwrapping algorithm from the general purpose computing environment of fringe-processing research into the real-time domain of signal processing applications. This application domain is characterised by the requirement to process data at a rate that keeps up with the system input, and the feasibility of implementations based on various hardware architectures are considered.

Chapter 4 explores various computationally efficient extensions to the generic temporal phase unwrapping algorithm that enable high-precision measurements of surface profile to be made using the method of white-light sinusoidal fringe projection. The development of a high-speed, high-accuracy profilometry system based on the temporal phase unwrapping algorithm and implemented using a pipeline image processing hardware architecture is described. An example data set measured using the system is presented.

Chapter 5 presents the results of a systematic experimental investigation into the issues affecting the performance of white-light projected fringe profilometers implemented using spatial light modulators.

Chapter 6 considers several extensions to the generic temporal phase unwrapping algorithm that are desirable for quantitative analysis of speckle interferograms. The chapter describes the design of a digital speckle pattern interferometry system based on the extended temporal phase unwrapping algorithm and implemented on a pipeline image processing hardware architecture. The system is able to display quasi-live two-dimensional pseudo-colour unwrapped phase maps computed in real-time from speckle interferograms acquired at normal frame rates.
Chapter 7 describes an experimental set-up for measuring out-of-plane surface displacement using the instrument described in chapter 6. Results are presented for several specimens under known loading conditions and are used to validate the unwrapping algorithm implementation. The ability to detect defects within composite materials is demonstrated using a custom inspection hood to load the specimen.

Chapter 8 presents the conclusions of this thesis, and provides a number of suggestions for future work.
Section I:
Literature review
1 Whole-field optical metrology

1.1 Introduction

Conventional methods of measuring physical parameters such as surface strain, displacement, and profile utilise strain gauges, dial gauges and other mechanical or electrical sensing devices. Although these point-wise methods can potentially produce high-precision measurements their drawbacks include the requirement for contact with the surface under test and the localised measurement area. In general, large numbers of separate measurements are required to build up an overall picture of the physical parameter, however, when the required number of sensors exceeds $10^2$-$10^3$ the cost typically becomes prohibitive. There has thus been much interest in developing whole-field (or full-field) non-contact metrology techniques that are able to provide measurements over large areas of the object surface at any one time to overcome the laborious procedure of point-wise measurement. Optical metrology methods such as speckle interferometry and optical profilometry provide an attractive solution for many applications and can provide whole-field information equivalent to more than $10^5$ independent point-wise sensors. Sophisticated digital cameras with high spatial resolution (i.e. number of pixels), high temporal resolution (i.e. frame rate), and high accuracy (i.e. number of bits) have in many applications replaced photographic plates, thus enabling optical techniques to be adopted more easily.

Whole-field optical metrology is a broad subject area encompassing the measurement of physical properties of smooth or rough objects that can be opaque or transparent. The scope of this chapter is limited, however, to engineering interest in surface displacement and profile measurement of opaque objects. Applications are introduced that provide a wide choice of sensitivities and dynamic range. The high sensitivity offered by interferometric techniques is well suited to displacement measurement where limited dynamic range is required and is introduced in section 1.2. In particular, speckle interferometry offers strong potential for integrated systems that can be used outside of the laboratory due to relatively low cost components, variable sensitivity, and relative ease with which it can be set-up and used.
Coherent optical methods are well suited for metrology applications requiring high precision but many techniques are only able to operate over a small dynamic range. In some applications, such as testing the flatness of diffuse surfaces, this does not pose a significant problem, however, there are other situations where an extended dynamic range is required. Whilst coherent interferometry may be used for surface profile measurement, incoherent optical methods can also provide very practical solutions and are particularly suitable for profiling applications where large surface height variations and discontinuities are encountered. These techniques, introduced in section 1.3, enable measurement of deep surfaces where dynamic range is regarded as being as important as measurement sensitivity.

1.2 Surface displacement measurement

This section presents the most popular whole-field optical methods for measuring the displacement and displacement-gradient of test surfaces. The Cartesian components of the displacement vector \( \mathbf{d}(P) \) at point \( P(x,y) \) on the test surface are denoted by \( u(x,y) \), \( v(x,y) \) and \( w(x,y) \), respectively.

1.2.1 Smooth wavefront interferometry

Coherent signal processing requires that information about both the amplitude and phase of wavefronts be measured. However, all practical light detectors respond only to light intensity and it is therefore necessary to convert the phase information to variations in intensity for measurement purposes. Interferometry is a standard technique used to accomplish this task. A second coherent wavefront of known amplitude and phase (the reference wave) is added to the unknown wavefront (the object wave); the intensity of the sum then depends on both the amplitude and phase of the original wavefront.

Smooth wavefront interferometry refers to those interferometry techniques based only on the “smooth” wavefronts produced when all the optical elements in a system, including the surface under test, have a negligible surface roughness when compared to the wavelength of light. By contrast, optical techniques involving rough surfaces
compared to the wavelength of light (e.g. holography and speckle interferometry) produce “speckle” wavefronts and are discussed later.

We begin our analysis by describing the spatio-temporal distribution of a linearly polarized plane harmonic wave with angular frequency $\omega = 2\pi v$ and initial phase $\varphi$ at the origin. If the wavefront is propagating in the direction of the wavevector $\mathbf{k}$ then the complex amplitude is given by

$$A(\mathbf{r},t) = a_0 e^{i(k \cdot \mathbf{r} - \omega t + \varphi)} \quad (1-1)$$

Removing the spatial dependence for clarity, the intensity, or irradiance, of a general wave is given by

$$I(t) = c \varepsilon \langle A(t) A^*(t) \rangle \quad (1-2)$$

where $c$ is the speed of the wave propagation, $\varepsilon$ is the electric permittivity of the transmission medium, $\langle \rangle$ is the time average operator, and $^*$ denotes the complex conjugate. Since we will only be concerned with relative intensities within the same medium it is useful to neglect the scaling constants and set

$$I(t) = \langle A(t) A^*(t) \rangle \quad (1-3)$$

A practical light sensor, however, cannot follow the frequency of light and therefore is unable to measure the instantaneous intensity. The sensor thus integrates over a finite measurement time $T_m$

$$I = \lim_{T_m \to \infty} \frac{1}{T_m} \int_{-T_m/2}^{T_m/2} A(t') A^*(t') dt' \quad (1-4)$$

Typically the integration period $T_m$ extends over many periods of oscillation (i.e. $T_m \gg 2\pi / \omega$) hence the measured intensity is simply

$$I = |A|^2 \quad (1-5)$$

Consider then an object wave $A_1 = a_1 e^{i\phi_1}$ and reference wave $A_2 = a_2 e^{i\phi_2}$, the intensity of the sum is given by

$$I = (A_1 + A_2)(A_1^* + A_2^*)$$

$$= a_1^2 + a_2^2 + 2a_1 a_2 \cos(\phi_1 - \phi_2) \quad (1-6)$$

The first two terms, the self-interference terms, depend only on the intensities of the two waves and represent noise terms that we have to suppress or discard as their
presence in the final analysis leads to poor signal-to-noise ratios. The third term, however, depends on their relative phases and thus encodes the phase information of the unknown object wavefront.

A useful alternative notation is

\[ I = I_0 + I_M \cos \Delta \phi \] (1-7)

where \( I_0 = I_1 + I_2 \) and \( I_M = 2\sqrt{I_1 I_2} \), where we define \( I_1 = A_1 A_1^* \), \( I_2 = A_2 A_2^* \), and \( \Delta \phi = \phi_1 - \phi_2 \).

A highly useful parameter in evaluating the performance of a system is the visibility or contrast of the interference pattern and is defined as

\[ V = \frac{I_{\text{max}} - I_{\text{min}}}{I_{\text{max}} + I_{\text{min}}} \] (1-8)

where \( I_{\text{max}} = I_0 + I_M \) and \( I_{\text{min}} = I_0 - I_M \). Hence, the intensity is often expressed in terms of visibility as

\[ I = I_0 (1 + V \cos \Delta \phi) \] (1-9)

### 1.2.2 Holographic interferometry

Advances in holography have resulted in two powerful interferometric techniques being developed that are suitable for non-destructive testing applications, and particularly measurements of the deformation of diffusely reflecting opaque objects. In contrast to a photograph where the light reflecting off the surface carries with it information about the irradiance but nothing about the phase of the wavefront that once emanated from the object, a hologram records both the irradiance and phase information. Following development of the holographic plate, the image is reconstructed by diffracting a coherent beam using the transparent hologram.

In the double exposure method of holographic interferometry two wavefronts scattered by the same object are mixed with a reference wavefront and recorded consecutively onto the same holographic plate. The plate is first exposed to the object wavefront scattered by the undisturbed object mixed with the reference wavefront (figure 1-1), then after a change in physical parameter (e.g. object shape due to deformation) the plate is exposed once again to the modified object wavefront mixed
with the reference wavefront (figure 1-2). Following development of the plate (figure 1-3), the reconstructed waves overlap and interfere to form a fringe pattern indicative of change in optical path length between the two object states. If the spatial variation of the interference phase over the observed reconstructed surface is low, the intensity distribution represents the irradiance of the object, modulated by a cosine fringe pattern as expressed in equation 1-9.

In real-time holographic interferometry the object under test is left in its original position throughout and only the object wavefront scattered by the undisturbed object mixed with the reference wavefront is recorded on the holographic plate. Following development of the plate it must be returned to its original position with sub-wavelength precision such that the reconstructed virtual image wavefront coincides exactly with the wavefront scattered directly by the object (figure 1-4). During testing of the object, superposition of the scattered wavefront with the reconstructed wavefront results in a live interference pattern (or interferogram) that is a function of the physical parameter under investigation. Video equipment is often used to make real-time recordings of the interferogram.

The displacement vector \( \mathbf{d}(P) \) of point \( P(x, y) \) on the object surface produces an optical path difference \( \delta l(P) \) given by[6]

\[
\delta l(P) = \mathbf{d}(P) \cdot [\mathbf{b}(P) - \mathbf{s}(P)]
\] (1-10)

where \( \mathbf{s}(P) \) and \( \mathbf{b}(P) \) are the unit vectors in the illumination direction and observation direction respectively. The interference phase is related to the path difference by

\[
\Delta \phi(P) = \frac{2\pi}{\lambda} \delta l(P)
\] (1-11)

so that we get the general formula

\[
\Delta \phi(P) = \mathbf{d}(P) \cdot \mathbf{e}(P)
\] (1-12)

where \( \mathbf{e}(P) \) is the so called sensitivity vector

\[
\mathbf{e}(P) = \frac{2\pi}{\lambda} [\mathbf{b}(P) - \mathbf{s}(P)]
\] (1-13)

In order to determine the components of a displacement vector from an evaluated interferogram we have to discriminate between a constant sensitivity vector and a
sensitivity vector that varies over the interferogram. In the latter case it may be sufficient to estimate the maximum error introduced by the extreme sensitivity vectors of the actual holographic set-up. In most engineering problems the deformation can be predicted and it is therefore good practice to configure the holographic set-up with maximum sensitivity in the expected direction of deformation. Abramson developed the so-called holo-diagram\cite{7} as an aid to designing holographic set-ups for optimum sensitivity.

Several excellent books\cite{8-10} are available providing further information on these holographic techniques, and many others.

### 1.2.3 Speckle interferometry

The granular appearance known as the speckle effect observed when coherent light is reflected from a diffuse surface can be a real practical nuisance in coherently illuminated systems. In holographic interferometry, for example, the speckle pattern is regarded as background noise. The development of speckle metrology\cite{11-13} is historically linked to the realization that the grainy appearance of the reconstructed holographic images could be used to convey additional information.

The simplest speckle metrology technique is speckle photography\cite{14} in which a single coherent beam is used to illuminate the object. Speckle patterns obtained before and after deformation are compared using the speckle grains recorded on the photographic plate as markers to monitor changes in the physical parameter. The principle sensitivity of speckle photography is in the direction normal to the optical axis (i.e. in-plane displacement). The displacements generally have to be larger than the speckle size, which can be controlled by the detector aperture. This is not strictly an interference method, however, the evaluation stage usually involves interference patterns.

Speckle interferometry\cite{8,15,16} involves the interference of two optical waves in which at least one is a speckle field. It provides a powerful technique for the measurement of deformations or contours of rough surfaces depending on the geometry of the illumination. Speckle interferometry is closely related to holographic interferometry and is equally sensitive. Whilst speckle interferometry does not require
the recording of a hologram, the fringe patterns obtained normally have a lower signal to noise ratio than those produced in holographic interferometry. Generally speaking, speckle interferometers can be divided into two categories, depending on whether two speckle fields (e.g. in-plane and shearing interferometry), or one speckle field and a smooth reference beam (e.g. out-of-plane interferometry), are mixed to form the interference pattern.

The most important technique of speckle interferometry is digital speckle pattern interferometry (DSPI), originally called electronic speckle pattern interferometry (ESPI), and also known by the names electronic holography and digital holography. Electronic speckle pattern interferometry[17] emerged in the 1970s as a method aimed at combining optics and electronics by using standard TV or CCD cameras rather than holographic or photographic plates to eliminate the slow and cumbersome chemical processing of photographic emulsions. Co-linear reference and object wavefronts are required together with an imaging system in order to produce interference patterns suitable for the relatively low-resolution cameras.

Speckle interferograms obtained by superposition of a reference beam do not contain fringes as in classical interferometry since the phase of each speckle is statistically independent. However, each speckle within the image plane of the detector can be considered as an independent interferogram that can be expressed using classical interferometry as in equation 1-7. The phase distribution in a speckle interferogram can therefore be described by a uniform randomly distributed variable over the range $[0, 2\pi]$. An in depth discussion of the statistics governing interferometry with speckle fields would be inappropriate in this thesis, however, a brief summary of the main conclusions which may be of interest for the practical optimisation of speckle interferometers are presented in appendix A1.

The measured physical parameter can be visualised by so-called correlation fringes obtained, for instance, by calculating the square of the difference in the intensity distributions (primary interferograms) recorded before and after deformation of the object. From equation 1-7 the intensity of each speckle grain (formed by the superposition of the object and reference beams) in the primary interferogram recorded with the object in its reference state is given by
Whole-field optical metrology

\[ I_i = I_0 + I_M \cos \phi \] (1-14)

where \( \phi \) is the random initial phase. Assuming that deformation of the object changes the phase but not the amplitude, the intensity of each speckle grain recorded in the deformed state is given by

\[ I_d = I_0 + I_M \cos(\phi + \Delta \phi) \] (1-15)

where \( \Delta \phi \) is the phase change due to deformation. The square of the difference is therefore

\[
(I_i - I_d)^2 = \left[ 2I_M \sin \left( \frac{\phi + \Delta \phi}{2} \right) \sin \frac{\Delta \phi}{2} \right]^2
\]

\[ = 2I_M^2 \sin^2 \left( \frac{\phi + \Delta \phi}{2} \right) (1 - \cos \Delta \phi) \] (1-16)

The low frequency \( \cos \Delta \phi \) term modulates the high frequency speckle noise to produce grainy correlation fringes. Typically these fringes represent contours of constant displacement component, or depth, depending on the optical arrangement used to acquire the speckle patterns. Dedicated systems have been commercially available for several years that use electronic subtraction and rectification to calculate and display the correlation fringes in real-time; such systems are well suited for qualitative analysis of object behaviour. Although intensity based techniques such as correlation fringes are sometimes also used for quantitative analysis, methods based on phase-shifting techniques are generally more appropriate and are discussed in detail in chapter 2.

There are many optical configurations used in speckle interferometry. Most of them are two-beam systems with each wavefront coming from a single light source and their paths separated by amplitude division. The most common arrangements for measuring surface displacement and displacement-gradient measurement are described in sections 1.2.4 to 1.2.6.

**1.2.4 Out-of-plane speckle interferometry**

Measurement of the out-of-plane displacement component \( w \) with speckle interferometers can be performed using either a speckle-wavefront or smooth-wavefront reference beam. The latter method has so far been the most popular and a
Whole-field optical metrology

common interferometer arrangement is shown in figure 1-5. The test surface is illuminated by the object beam parallel to the x-z plane and imaged onto the detector surface through a lens arrangement. The smooth-wavefront reference beam is brought onto the optical-axis using the beam combiner and superimposed with the speckle image of the object at the detector.

There have been many variations on this scheme to reduce the secondary reflections of the reference wave. For example, wedge[18] and cube beamsplitters have been used to introduce the reference beam, as well as monomodal optical fibre. In addition many arrangements introduce the reference beam before the imaging lens, as demonstrated in the interferometer arrangements described in sections 6.4.2 and 7.2.2, for example.

From equation 1-13, the sensitivity vector e(P) at point P in plane x-z on the test surface is given by equation 1-17, where θ is the object beam angle of incidence to the surface normal.

\[
e(P) = \frac{2\pi}{\lambda} \begin{bmatrix} \sin \theta \\ 0 \\ 1 + \cos \theta \end{bmatrix}
\]

(1-17)

In the plane x-z, the interferometer has zero sensitivity to the in-plane displacement component v, with negligible sensitivity above and below this plane for small regions of interest. For general operation, it is desirable to minimise the sensitivity to in-plane displacement component u by keeping θ as small as possible. Hence, for an illumination beam at normal incidence, adjacent coherence fringes correspond to an out-of-plane displacement of \(\lambda/2\). The distance between the test surface and the effective source of the object beam should be maximised so that variations in \(\theta\), and thus variations in the sensitivity vector, are minimised across the region of interest. It should be noted that plane wavefronts would be required to achieve completely uniform sensitivity vectors, however, for many practical applications this approximation is acceptable. Wan Abdullah et al.[19] provide a detailed analysis of the errors introduced by this approximation.

High sensitivity to out-of-plane displacement can be a hindrance when measuring large deformations since high fringe densities prevent good quantitative spatial
analysis of the interferogram. For example, the upper limit for a $512 \times 512$ pixel CCD
detector is approximately 20 fringes thus limiting the maximum deformation that can
be reliably measured to approximately $7 \mu m$. Several methods for reducing the
sensitivity have been reported, including for example, two-beam illumination with a
small angle between the beams[20], oblique incidence and observation[21], and
longer-wavelength lasers[22].

1.2.5 In-plane speckle interferometry
The double-illumination system[23] shown in figure 1-6 for measuring in-plane
displacement components uses two coherent beams of light at angles $\theta_1$ and $\theta_2$ either
side of the normal. The two beams generate their own speckle patterns that combine
by superposition, and the scattered light is imaged on to the detector. It can be shown
that the resultant sensitivity vector for such double-illumination arrangements is given
by

$$e(P) = \frac{2\pi}{\lambda} \left[ s_x(P) - s_y(P) \right]$$

(1-18)

where $s_x(P)$ and $s_y(P)$ are the unit vectors along the illumination directions. The
sensitivity vector at point $P$ in plane x-z on the test surface is therefore given by

$$e(P) = \frac{2\pi}{\lambda} \begin{bmatrix} \sin \theta_1 - \sin(-\theta_2) \\ 0 \\ \cos \theta_1 - \cos(-\theta_2) \end{bmatrix}$$

(1-19)

As with the optical arrangement discussed in section 1.2.4, there is negligible
sensitivity to in-plane displacement component $v$. Out-of-plane sensitivity is zero at
all points on the object surface where the illumination beams are symmetrical about
the normal ($\theta = \theta_1 = \theta_2$) and the in-plane sensitivity component simplifies to

$$e_z = \frac{4\pi}{\lambda} \sin \theta$$

(1-20)

The presence of divergent illumination beams results in small variations in the
sensitivity vector across the region of interest and is minimised by increasing the
distance between the effective source position of the illumination beams and test
surface. The presence of the $\sin \theta$ term in equation 1-20 enables the sensitivity to in-
plane displacement $u$ to be changed by varying the illumination direction.
Correlation fringes indicate contours of equal in-plane displacement parallel to the plane containing the two illumination beams. Hence, measurement of in-plane displacement component \( v \) along the \( y \)-axis would require that the illumination beams be arranged parallel to the \( y-z \) plane. In general, fringe quality is lower to that of a uniform reference system due to the speckle statistics. Since speckle decorrelation limits the displacement range that may be measured, large in-plane displacements must be measured by integrating incremental speckle displacements for which the decorrelation remains small.

### 1.2.6 Speckle shearing interferometry

Electronic speckle-shearing pattern interferometry (ESSPI) was first suggested by Leendertz et al. [24] and has evolved into an essential tool for measuring the partial derivative of displacement of rough surfaces. A simple speckle shear arrangement is shown in figure 1-7 that uses a single illumination beam and a Michelson interferometer to perform shearing of the wavefront. A small tilt in the angle of one of the mirrors produces a constant linear shear \( \Delta x \) resulting in two superimposed images of the object on the detector that combine coherently to generate a third speckle pattern. Each point in the image plane therefore receives a contribution from two points \( P_1(x, y) \) and \( P_2(x + \Delta x, y) \) on the test surface, where the object plane shear \( \Delta x \) is related to the image plane shear \( \Delta x \) through the magnification of the imaging lens. One of these mutually shifted images can be regarded as the reference wavefront and the other the object wavefront, and hence this approach is known as self-reference.

As in the previous methods, the image captured after deformation is subtracted from the reference image captured before deformation. From equation 1-12, the phase change in the contributions from points \( P_1 \) and \( P_2 \) due to the deformation of the test surface are given by equations 1-21 and 1-22, respectively.

\[
\delta \phi_1 = d(P_1) \cdot e(P_1) \quad (1-21)
\]

\[
\delta \phi_2 = d(P_2) \cdot e(P_2) \quad (1-22)
\]
If we let the Cartesian components of the displacement vectors $d(P_1)$ and $d(P_2)$ be $(u, v, w)$ and $(u + \Delta u, v + \Delta v, w + \Delta w)$, respectively, then the interference phase of the combined speckle patterns is $\Delta \phi = \delta \phi_1 - \delta \phi_2$ is given by

$$\Delta \phi = \frac{2\pi}{\lambda} \left[ \Delta w(1 + \cos \theta) - \Delta u \sin \theta \right]$$  \hspace{1cm} (1-23)

Dependence on the in-plane displacement component $u$ is dropped if the illumination beam is arranged normal to the object surface, i.e. $\theta = 0$. Hence, provided that the object plane shear $\Delta x$ is small compared to the surface displacement then the change in phase can be expressed in terms of the partial derivative of out-of-plane displacement with respect to the $x$-direction.

$$\Delta \phi \approx \frac{2\pi}{\lambda} \left[ (1 + \cos \theta) \frac{\partial w}{\partial x} \right] \Delta x$$  \hspace{1cm} (1-24)

The shear interferogram displays only abrupt changes in displacement as contributions due to rigid body movement of the test surface cancel. Alternative methods exist to create the sheared image and include: (a) the use of a glass wedge placed in front of one half of the imaging lens[25]; (b) the use of two tilted glass plates[26] placed in front of the imaging lens; (c) the use of a binary grating[27] placed in front of the imaging lens; (d) the use of a double wedge or Fresnel bi-prism[28] placed in front of the imaging lens; and (e) the use of multiple apertures in conjunction with a lens and defocusing[29,30]. Krishna Murphy et al. [31] describe some of the drawbacks of these techniques and introduces the split-lens shearing method as an alternative.

Due to the common beam path, the shearing interferometer provides a simple arrangement with good immunity to the effects of turbulence and vibration for small shear amounts, and therefore is well suited to industrial applications. The shearing speckle interferometer is commonly employed for detecting defects in surfaces, where the defect produces a steep out-of-plane displacement gradient identified by the close grouping of fringes. Increasing the shear distance improves gradient sensitivity, whilst at the same time reducing spatial resolution. Significant rigid body movement or turbulence may result in speckle decorrelation reducing fringe contrast to unacceptable levels.
1.2.7 Difference techniques

The interferometer arrangements introduced in the preceding sections provide important techniques for measuring surface displacement of the order $10^{-6}$ m, with sensitivity comparable to the wavelength of light. However, this very high sensitivity is sometimes undesirable, particularly when measuring large surface displacements where the high fringe densities present in the resulting interferogram prevent reliable extraction of quantitative data. Many engineering problems produce displacements of the order $10^{-3}$ m and above thus rendering these interferometer techniques impractical. In addition, since these methods are coherent processing techniques they are susceptible to small amplitude vibrations and air turbulence that are often present in industrial environments.

In such circumstances it may therefore be more appropriate to calculate the surface displacement from the difference in surface profile before and after loading. The following sections describe several whole-field optical surface profiling techniques that provide a range of sensitivities in the z-direction. Furthermore, several of these methods use incoherent illumination and are therefore less susceptible to air turbulence.

1.3 Surface profile measurement

1.3.1 Introduction

Techniques presented in this section have been developed for the three dimensional description of deep surfaces and their position in space. Most surface profiling systems trade off measurement accuracy for dynamic range and in general are less sensitive than the interferometry techniques presented in the previous section. The methods described here, for example, have measurement sensitivities ranging from $10^{-6}$ m to $10^{-2}$ m. However, there is a recognised requirement in industry for optical techniques than provide both high dynamic range and high precision. Applications for these systems include automated manufacturing, quality control, robotic vision, and solid modelling[32].

Non-contact measurement of surface profile is usually dependent on techniques based on image cues, triangulation, various interferometric methods (including wavelength
Whole-field optical metrology

change, displacement of the test surface, and shifting the illumination beams) and projection of structured light patterns. These techniques can be divided into the broad categories of passive or active sensors. Passive profile sensors are typically less invasive than active methods and remotely measure the test surface profile under natural illumination by examining image cues such as shading or texture. By contrast, active methods typically require temporal control of the illumination, focus, or relative position of the test surface, and purposively adjust this parameter during the measurement process to increase the dynamic range whilst maintaining measurement accuracy.

Surface profilometry has been evolving as a subject for about thirty years and compiling an exhaustive list of all of the available methods is difficult to collate due to its vast, multidisciplinary, and continuous expansion. There is more to profilometry than can be covered in this chapter, and therefore only the most widely used whole-field techniques that are suitable for automation are introduced. This chapter does not discuss whole-field optical profilometry methods based on physical scanning of point-wise detectors, such as optical laser scanners for example, and the reader may refer to a recent paper Amann et al.[33] that provides a critical review of such techniques. In addition, Chen et al.[32] have recently published an overview of whole-field optical metrology methods for shape measurement.

The following discussion begins with an introduction to computer vision techniques for shape measurement. In practice these methods tend to have quasi real-time constraints and are suitable only for obtaining broad characterisation of surface properties. The remaining range imaging methods described in this chapter have been developed to determine surface profile with accuracy several orders greater than that attainable by computer vision methodology, and furthermore, may also be suitable for quasi real-time applications. In general, the range imaging methods vary one or more of the system parameters over the measurement period to extend dynamic range whilst maintaining high accuracy.

Profiling of complex objects requires registration from different views, resulting in point clouds related to different co-ordinate systems. Precise transformation of these point clouds from each of the individual sensor co-ordinate systems into a common
global co-ordinate system is therefore required to accurately characterise the complete surface. Several approaches are currently used to determine the relative orientation of the sensors including: (a) positioning using high-precision mechanical actuators and transducers (for example, the FaroArm manufactured by Faro Technologies Inc. enables multi-axis positioning), (b) statistically matching of point cloud data by optimal fitting[34], and (c) photogrammetric matching of point cloud data[35].

1.3.2 Computer vision techniques
Several techniques for determining surface profile fall into the domain of computer vision (also known as scene analysis or image understanding). In general, it is not possible to obtain depth information directly from two-dimensional intensity images, and it is therefore necessary to invert the many-to-one imaging transformation that maps points in the scene on to an image plane. In general, this recovery process requires knowledge of the objects in the scene and the projection geometry.

The shape-from-texture[36] method exploits variation in the size, shape, density, and aspect ratio of texture primitives to provide clues for the estimation of surface orientation and ultimately surface profile of scene objects. For successful measurement, the detector must be able to resolve the surface texture such that texture primitives can be distinguished. In practice, however, automated delineation of the primitives is extremely difficult for complex objects and the technique is thus only suitable for simple surfaces, or surfaces for which the texture is well characterised.

Exploitation of variations in image intensity due to changes in surface slope and a-priori knowledge of the reflectance map can be used to determine the surface position and orientation. The reflectance map specifies the intensity of surface patches at particular orientations for a given distribution of illumination and surface material. The basic technique, known as shape from shading, uses a single light source, however, unambiguous solutions are generally only obtained when the number of degrees of freedom can be reduced by assuming that the scene is composed of diffuse, smooth, continuous surface patches. Photometric stereo extends the technique by using several light sources. When a diffuse surface is illuminated by a point source, the lines of constant intensity in the image can be described by second-order polynomials. The set of surface orientations that can generate a given polynomial are
restricted to those that lie along a curve in the reflectance map. A minimum of three independent point sources is therefore required such that the intersection of these curves unambiguously defines the surface orientation.

Another popular technique is stereo vision (also known as binocular vision), in which depth information is extracted from a pair of images obtained using two cameras displaced from each other by a known baseline distance. The simplest model uses two identical cameras arranged so that their image planes are coplanar. A feature in the scene is viewed by the two cameras at different positions in the image plane, separated by a disparity vector. Depth information can be recovered by identifying the disparities of corresponding image points (known as homologous points). Increasing the baseline distance enhances the accuracy of the depth calculation, however, it may also introduce other errors (e.g. distortions introduced by the perspective projection). There has been much research[37] applied to the detection and matching of features in the image pairs using methods such as edge matching, region correlation, and multiple primitives. This correspondence problem remains the main factor limiting adoption of the technique.

A further method known as shape-from-focus exploits the finite depth of field of optical systems to recover depth information[38]. The image is modelled as a convolution of focused images with a point spread function determined by the camera parameters and distance of the object from the camera. However, such reconstruction problems are mathematically ill posed, and height information can only be recovered in those regions where surface features are present.

1.3.3 Photogrammetry

Crossing over from the discipline of computer vision, photogrammetry techniques for determining three-dimensional information pursues higher levels of reliability and accuracy. Photogrammetry typically extends, but is not restricted to, the stereo vision method and developed almost exclusively for mapping large geographic areas in the domain of civil engineering and remote sensing. The advent of low cost digital cameras and software for automated image analysis has made the technique more attractive for application to close range measurement problems.
The starting point for building a close range photogrammetry model is the *central perspective projection* as shown in figure 1-8. The primary Cartesian co-ordinate system \((XYZ)\) is located arbitrarily in object space and the secondary co-ordinate system \((xyz)\) has its origin at \(O\); its z-axis coincides with the principal axis \(POp\) and is directed away from the projection plane; its \(x\)- and \(y\)-axes are parallel to the projection plane. A point \(A(X_A, Y_A, Z_A)\) in the three-dimensional object space is projected to point \(a(x_a, y_a, -c)\) on the projection plane by a straight line \(AOa\) passing through the *perspective centre* \(O\). The distance \(Op\) from the perspective centre to the projection plane is the *principal distance* denoted by \(c\). The transformation mapping primary co-ordinates to secondary co-ordinates is given by the *collinearity equations*,

\[
x_a = \frac{-c[r_{11}(X_O - X_A) + r_{12}(Y_O - Y_A) + r_{13}(Z_O - Z_A)]}{r_{31}(X_O - X_A) + r_{32}(Y_O - Y_A) + r_{33}(Z_O - Z_A)}
\]

\[
y_a = \frac{-c[r_{21}(X_O - X_A) + r_{22}(Y_O - Y_A) + r_{23}(Z_O - Z_A)]}{r_{31}(X_O - X_A) + r_{32}(Y_O - Y_A) + r_{33}(Z_O - Z_A)}
\]

(1-25)

where \(r_{ij}\) are the elements of the rotation matrix \(R\) that maps the primary axes onto the secondary axes. The object space co-ordinates \((X_A, Y_A, Z_A)\) of target \(A\) cannot be determined unambiguously from the photo co-ordinates \((x_a, y_a)\) since the reverse transformation does not provide a unique solution. A measurement is analysed from at least two positions, yielding four measurement values; the redundant value can be used for both instant determination of object co-ordinates and model parameters. In general, three-dimensional reconstruction is based on the *bundle adjustment* principle.

A geometric model of the camera positions and scene co-ordinates is developed analytically from the orientation of bundles of light rays and the optimal solution is found using a least squares minimisation approach. To determine the scale, only a single distance in the object space must be known. The technique enables *self-checking* of the data quality and *self-calibration* (i.e. simultaneous determination of co-ordinates and system parameters) to be performed, removing the need for expensive standards as used in conventional calibration procedures. Accuracies as high as 1 part in 1,000,000 have been reported[39], however, calculations are often time-consuming. In general markers must be attached to the test surface as homologous points to aid in solving the correspondence problem unambiguously. Recent work by Reich et al.[35] has combined photogrammetry and fringe projection.
to solve the correspondence problem by projecting coded fringes with different orientations onto the test surface.

An excellent introduction to close range photogrammetry is provided by Atkinson[37] and there are several books which cover the general technique[40,41].

1.3.4 Contouring speckle interferometry

Although speckle interferometry is most often encountered in the domain of surface displacement analysis, several methods based on speckle interferometry have been developed for measuring surface profile. These methods generally vary one of a number of parameters controlling the fringe formation against time, including: (a) illumination direction (described below); (b) temporal coherence (for example, see section 1.3.5); and (c) optical wavelength (for examples, see sections 1.3.6 and 1.3.7).

The electronic speckle contouring (ESC) method is based on conventional in-plane and out-of-plane displacement sensitive arrangements where small changes in the illumination direction are introduced. Rodriguez-Vera provides useful state-of-the-art review of ESC methods in reference [42]. Zou et al.[43] and Rodriguez-Vera et al.[44] describe a dual-beam arrangement for measuring surface profile in which the surface remains unchanged and the illumination beams are shifted to vary the sensitivity vector during the experiment. Joenathan et al.[45] describe an analogous arrangement in which the object is tilted to induce a relative shift in the illumination directions. The latter method uses an arrangement that is similar to the in-plane displacement set-up (figure 1-6), with the addition of a rotation stage on which the test surface is mounted. The illumination beams are arranged symmetrically about the normal ($\theta = \theta_1 = \theta_2$) such that the interferometer is only sensitive to the in-plane displacement component $u$ (see equation 1-20) as discussed in section 1.2.5. Correlation fringes are formed from the speckle interferograms recorded before and after a small rotation of the surface. The in-plane displacement component $u$ produced by a rotation $\omega$ is dependent on the surface height $h$, and therefore the resultant fringes represent surface height contours.

Consider the arrangement shown in figure 1-9(a). The surface is rotated about an axis parallel to the $y$-direction and passing through point $O$. The so-called tilt plane $TP$ is
defined as the plane passing through the axis of rotation in which the in-plane displacement is zero (i.e. perpendicular to the line of sight). The point $P(x, y, z)$ on the surface, located at height $h$ in front of the tilt plane, is displaced by the rotation to $P'(x + \Delta x, y, z + \Delta z)$. Geometry analysis (see figure 1-9(b)) reveals that the in-plane displacement component $u(P)$ is given by

$$u(P) = 2r \sin(\omega / 2) \cos(\alpha - \omega / 2)$$  \hspace{1cm} (1-27)

where $r$ is the radial distance between $O$ and $P$. Substituting $r = \frac{h}{\cos \alpha}$ gives

$$u(P) = \frac{2h \sin(\omega / 2) \cos(\alpha - \omega / 2)}{\cos \alpha}$$  \hspace{1cm} (1-28)

For small angles of rotation $\omega$, the in-plane displacement is approximately proportional to the surface height and angle of tilt

$$u(P) \approx h \sin \omega$$  \hspace{1cm} (1-29)

Substituting equations 1-20 and 1-29 into equation 1-12 gives

$$\Delta \phi(P) = \frac{4\pi}{\lambda} h \sin \omega \sin \theta$$  \hspace{1cm} (1-30)

where $\Delta \phi(P)$ is the change in speckle phase contributed by point $P$, due to rotation of the test surface. Coherence fringes thus represent surface height contours separated by $h = \frac{\lambda}{(2 \sin \omega \sin \theta)}$, where the dependence on $\theta$ enables the height sensitivity to be controlled simply by changing the illumination directions. Fringe modulation is maximised by reducing speckle decorrelation and therefore small tilt angles should be used. Height contour sensitivity can be enhanced by a factor of $N$ simply by modulo-$2\pi$ addition of the phase changes due to $N$ small tilt increments.

The close relationship between speckle interferometry and holography means that the same principles can be extended to the holography domain, in which the chosen parameter is altered between constructions of the two wavefronts. Holographic optical arrangements have been developed for contouring by shifting the illumination direction[46-50], by altering the refractive index of the surrounding medium[51] (also called the immersion method), and by wavelength differences[52-59].
1.3.5 Low coherence interferometry

Temporal coherence describes the correlation of a wave with itself as a function of distance along the direction of propagation. Although the degree of coherence cannot be measured directly, it can be determined from the fringe contrast, which is measurable. For waves of equal intensity, the degree of coherence is numerically equal to the contrast (equation 1-8), in other cases additional terms are introduced[60].

The coherence time $\tau_c$ is defined as the time shift at which the contrast falls to $1/e$.

In interferometers, for example, each arm has a different optical path length and therefore introduces a time shift between the two light waves and it is useful to define the coherence length $l_c = c \tau_c$.

Dresel et al.[61] described a technique known as coherence radar (also called low-coherence interferometry) that exploits the coherence length of a given light source to measure surface profile. The technique is based on the generation of white-light fringes by nulling the optical path differences caused by the surface height distribution. The arrangement (figure 1-10) is based on a Michelson interferometer with one of the mirrors replaced by the test surface that is mounted on a mechanical translation stage. Mirror $M$ is mounted on a piezoelectric transducer (PZT) to enable phase shifting. A light source with a short coherence length is required for accurate measurement, typically a laser diode, a LED, or incandescent lamp. The short coherence length restricts interference to those speckles that correspond to the surface elements close to the plane $R$ where the optical path lengths of the interferometer arms balance. The intensities of the interfering waves must be equal for maximum contrast and therefore a neutral density filter is introduced into the reference arm to compensate for scattering at the test surface. The imaging lens is adjusted so that plane $R$ is focused onto the detector.

During the measurement process the test surface is slowly translated along the $z$-axis. For a given offset, $h(t)$, along the $z$-axis, the reference arm is phase stepped using the PZT and the contrast $I'(x, y, t)$ calculated independently for each pixel from the measured intensities. The surface height at each pixel is then determined by detecting the offset $h_{\text{max}}(x, y)$ required for maximum contrast at each pixel. It should be noted
that this method does not measure the phase of the interference, but merely identifies that coherent interference is occurring.

Since the illumination and observation directions are parallel the technique does not suffer from shading problems. In addition, the measurement accuracy is independent of aperture setting (unlike shape-from-focus methods for example) and measurement distance so the technique is well suited to surfaces with deep narrow holes. For practical broadband light sources the illumination aperture should be smaller than the observation aperture to maintain spatial coherence. Accurate mechanical movement of the object and reference mirror is required, and interferometer alignment may become a critical issue for large translations. The measurement process can be time consuming: a total translation of 5 cm, for example, by steps smaller than the coherence length (e.g. 2 µm) requires a repetition of more than 25,000 step movements of the test surface, with phase shift measurements made at each step.

1.3.6 Fourier transform speckle profilometry

Fourier transform speckle profilometry (FTSP) suggested by Takeda et al.[62] is based on the combination of wavelength-shift speckle interferometry[63] and the Fourier-transform technique (FTT) for temporal fringe pattern analysis (refer to section 2.2.6). The technique overcomes some of the shortcomings of coherence radar since it does not depend on the accurate mechanical translation of the surface and reference mirror, thus reducing hardware costs and measurement time. The optical arrangement (figure 1-11) is similar to that used for coherence-radar but uses a coherent frequency-tuneable laser diode as the light source. The test surface is placed in one of the interferometer arms such that the half optical path difference \( l(x, y) \) depends on the surface height. The basic technique scans the laser diode injection current so that the wave number \( k(t) = \alpha t + \kappa(t) \) varies quasi-linearly with time, where \( \alpha \) is a constant and \( \kappa(t) \) represents an initial wave number \( k(0) \) plus an unavoidable deviation from perfect linearity. Wavelength scanning causes the intensity of each speckle to vary sinusoidally giving the illusion that the speckle pattern is “boiling”: the rate of intensity variation is proportional to the optical path difference. The time-varying specklegram is of the form
Whole-field optical metrology

\[ g(x, y, t) = a(x, y, t) + b(x, y, t) \cos[2k(t)l(x, y)] \]  
(1-31)

where \( a(x, y, t) \) and \( b(x, y, t) \) are, respectively, the background intensity and the fringe amplitude, that unavoidably change with the injection current. Substituting for \( k(t) \) gives equation 1-32 that describes an amplitude- and phase-modulated sinusoid with a temporal carrier frequency \( f_0(x, y) \),

\[ g(x, y, t) = a(x, y, t) + b(x, y, t) \cos[2\pi f_0(x, y)t + \phi(x, y, t)] \]  
(1-32)

where

\[ f_0(x, y) = \frac{a(x, y)}{\pi} \]  
(1-33)

\[ \phi(x, y, t) = 2\kappa(t)l(x, y) \]  
(1-34)

The FTT provides a robust method for extracting the height information encoded in the temporal carrier frequency since it is immune to non-linearity of the injection current vs. wavelength characteristic and variations due to temperature. If we let \( c(x, y, t) = \frac{1}{2} b(x, y, t) \exp[\phi(x, y, t)] \) then the Fourier transform of \( g(x, y, t) \) with respect to the time variable \( t \) is given by

\[ G(x, y, f) = A(x, y, f) + C[x, y, f - f_0(x, y)] + C^*[x, y, f + f_0(x, y)] \]  
(1-35)

where \( \cdot \) denotes the complex conjugate, the uppercase letters denote the Fourier spectra of the signals denoted by the corresponding lower case letters. Since the separation of these spectra is proportional to \( l(x, y) \) it is necessary to arrange the interferometer with a large optical path difference between the arms to enable the spectra to be separated. One consequence of this increased path difference, however, is that the temporal sampling rate must also be increased to avoid aliasing problems. One of the sidebands is selected by filtering and the inverse Fourier transform computed to obtain the analytical signal \( c(x, y, t) \). The complex logarithm is calculated to separate the amplitude- and phase-modulation terms. Finally, one-dimensional phase unwrapping along the time axis is required to recover \( \phi(x, y, t) \), and thus obtain the height information.

As with the coherence radar technique, Fourier transform speckle profilometry is a line-of-sight measurement method and therefore does not suffer problems due to shading. However, some of the light scattered back from the surface of the object will
not enter the interferometer and therefore some pixels may not contain valid profile information and must be masked from the result.

1.3.7 Spectral interference microscope
Commercial tuneable laser diodes tend to have many mode hops along the frequency axis that limit the continuous frequency scanning range available to FTSP systems and so restrict the achievable vertical measurement resolution to approximately \( 10^{-4} \) m. Attempts have been made to improve the resolution by use of more sophisticated lasers, such as an external cavity laser diode[64,65] and a dye laser[66,67] that have a wider tuneable range. However, these lasers are often expensive, and in the case of the dye laser relatively cumbersome due to the use of carcinogenic chemicals and water cooling.

The spectral interference microscope developed by Kinoshita et al.[68] provides an alternative solution by exploiting a frequency tuneable liquid crystal Fabry-Perot etalon device that acts as a very narrow band-pass filter. The device has been developed primarily for wavelength multiplexing communication applications and consists of a crystal with a partially mirrored surface on each end and transparent electrodes. An applied voltage in the range 3-5 volts changes the orientation of the molecules within the crystal and so controls the refractive index. As a direct result, the applied voltage can be used to conveniently control the pass-band of the Fabry-Perot filter.

Liquid crystal Fabry-Perot interferometers (LC-FPI) can be used together with inexpensive broadband light sources (e.g. LEDs and super luminescent diodes) to select the wavelength of interest. Reference [68] describes a LC-FPI used as an optical frequency scan device and designed with a nominal pass-band width of 1.1 nm at the central wavelength of 665 nm that can be varied over a nominal maximum range of 23 nm. The system is able to perform measurement on discontinuous microscopic surfaces without the use of mechanically moving components. In addition, the LC-FPI is physically very small and only requires low voltages to control the pass-band.
1.3.8 Coaxial coimage plane projection and observation profilometry

Interferometry contouring methods enable surface profiles to be measured without the baseline required for triangulation systems and therefore do not suffer from shading problems. However, these coherent processing systems are susceptible to vibration noise and air currents, which makes them unsuitable for adverse industrial environments. The coaxial coimage plane projection and observation technique proposed by Takeda et al. [69] combines the robustness of white light projection with the immunity to shading problems. The arrangement shown in figure 1-12 consists of a projection system that projects a grating pattern onto the conjugate image plane $R$, and an observation system that images plane $R$ via a beam splitter onto the detector. Plane $R$ is therefore a coimage plane for both the projection and observation systems that are arranged coaxially to eliminate shading problems. As the test surface is moved through the conjugate plane $R$ using a high precision translation stage the contrast of the detected fringe pattern varies, with the peak contrast occurring where the scattering surface crosses the plane $R$. The surface height is therefore determined by detecting the translation required for maximum contrast at each pixel.

This arrangement can be regarded as a depth-from-focus technique, however, it differs from conventional depth-from-focus methods that rely on the spatial information of the surface itself rather than projected spatial information, and exclusively defocus the observation optics. The use of projected grating patterns enables profiling of non-textured surfaces and, furthermore, the reduced depth of focus of the combined imaging system provides an improvement in the measured height resolution. One of the main advantages of this arrangement over projection systems that require a finite baseline is that problems due to shading are significantly reduced. Note, however, that this is arrangement cannot be regarded as a “true” line-of-sight technique since a finite range of angles along the coaxial projection and observation direction are required to provide a finite depth-of-field.

1.3.9 Binary coding of projected structured light

Interferometric techniques discussed in the previous sections are able to achieve very accurate measurements over small depth ranges but are unsuitable for medium- and long-range applications. *Active structured light* methods based on triangulation (also
Whole-field optical metrology

categorised as active triangulation methods) are widely used in this interval and use a projection unit to illuminate the scene with a known light pattern coming from a known angle relative to the viewing angle.

In locating a point \( P(X, Y, Z) \) on the test surface, the camera pixel co-ordinates \((x, y)\) determine two of the three degrees of freedom and the projector is used to label the third. To determine surface height it is sufficient to label planes in space and this is achieved by projecting structured light patterns onto the test surface. The projection of lines can be regarded as light sectioning planes that label the measurement volume as illustrated in figure 1-13. In general, the sections are labelled by the intensity of the illumination, however, other methods have been reported (see reference [70], for example). The number of sections that can be labelled by a given projection in a single image is therefore limited by number of different intensities of light that can be generated. Projection of a temporal sequence of intensities for each lighting section enables the dynamic range to be extended. The sequence of intensities can be regarded as a code that uniquely identifies a given section. In addition to the measured intensity values, parameters describing the geometrical conditions of the sensor enter into the co-ordinate calculation. These parameters are normally determined in a calibration procedure before actual three-dimensional measurement begins.

The choice of coding scheme influences the speed and resolution of the measurement and several approaches have been reported (see references [71,72] for example); the most common schemes employ \( n \)-bit binary codes (i.e. the projected stripe is either "on" or "off") and allow the unique description of \( 2^n \) different light sections. An optimum coding scheme should: (a) be self normalising to simplify demodulation; (b) identify each light section uniquely; (c) only change one of the code digits between adjacent code-words (i.e. Hamming-distance equal to 1) to avoid decoding errors; and (d) utilise as many of the available code-words as possible to maximise efficiency. Several alternatives schemes, including the popular Gray-code and Lemming-code, are evaluated by Gartner et al.[73].

Binary coding of projected structured light provides a convenient technique for automation and several systems have been developed using liquid crystal spatial light modulators to generate the projection sequence. Increasing the number of codewords
enables either the dynamic range or the resolution of measurement to be extended, however, this also increases the number of patterns that must be projected. The image analysis process can be regarded as a simple fringe counting method and as such the resolution is limited to integer fringe orders. Sansoni et al. [74] reported a technique to address this problem using the spatial phase shifting method to measure fractional parts of a code word; the technique has been adopted by most commercial profilometry systems based on binary coding available today. In general binary coding schemes are only practical for incoherent fringe projection systems and require high stability of the optical set-up to maintain pixel registration during the measurement period. As with all triangulation techniques a finite baseline is required between the projector and the camera that may introduce shading problems in the acquired images. Furthermore, measurement of dynamic events requires expensive high-speed projection and imaging hardware.

1.3.10 Spatial frequency scanning of projected structured light
Projected structured light techniques using projected gratings with sinusoidal intensity distribution are based on the extraction of the continuous phase parameter to identify the projection plane (see figure 1-14). The achievable measurement resolution is therefore not restricted by a finite number of codewords as seen with the binary coding approach. When a grating pattern is projected at an angle to the observation direction on to a surface the fringes are perturbed according to the topography and can be regarded as a phase-modulated pattern with a constant spatial carrier. Demodulation of the deformed grating by means of a matched reference grating results in the well-known moiré fringe pattern [75] and enables very accurate measurement of surface profile [76]. Several automated methods for quantitative analysis of the moiré fringes have been described [77,78], however, the sign of the depth information cannot be determined unambiguously. Xie et al. [79,80] describe a simple extension to the shadow moiré technique that enables absolute height information to be determined by rotation of the grating. Several alternative approaches to moiré contouring based on analysis the deformed grating itself rather than the moiré pattern have been developed [81] and therefore determine surface height unambiguously [62,82,83].
Considering the general non-telecentric divergent projection geometry shown in figure 1-15, point \( P \) is the centre of the exit pupil of the projector lens and point \( I \) is the centre of the entrance pupil of the imaging lens separated by distance \( d_0 \). The crossed optical axes of the projection and imaging lenses lie in the same plane and intersect making an angle \( \theta \) at point \( O \) on a fictitious plane \( R \). Plane \( R \), normal to the optical axis of the imaging lens, serves as a reference from which surface height \( h(x, y) \) is measured. We define the co-ordinate system \((x, y)\) lying in the fictitious plane \( R \) with the origin at \( O \). If the uniform grating image falling on to the plane \( Q \) perpendicular to the projection axis has a constant spatial frequency \( f = \lambda^{-1} \), then the fringe period of the grating image \( g_o(x, y) \) viewed on plane \( R \) will increase as point \( C \) on the \( x \)-axis departs from the origin. Grating intensity \( g_o(x, y) \) is described by a spatial carrier frequency \( f_0 \) phase-modulated by \( \varphi_o(x) \),

\[
g_o(x) = a_o(x, y) + b_o(x, y) \cos[2\pi f_0 x + \varphi_o(x)]
\]

where \( f_0 = \lambda^{-1} = \lambda^{-1} \cos \theta \) and \( \varphi_o(x) = 2\pi f_0 BC \) equals the phase difference along length \( BC \). To evaluate \( BC \) we note that \( BC = OC - OB \) and let \( OC = x \). From the geometry we have \( OB = l_0 \tan \alpha \cos^2 \theta \) where \( \tan \alpha = \left( x \cos^2 \theta \right) \left( l_0 + x \sin \theta \cos \theta \right)^{-1} \).

Substituting \( OB \) and \( OC \) gives \( BC = \left( x^2 \sin \theta \cos \theta \right) \left( l_0 + x \sin \theta \cos \theta \right)^{-1} \) and hence,

\[
\varphi_o(x) = \frac{2\pi f_0 x^2 \sin \theta \cos \theta}{l_0 + x \sin \theta \cos \theta}
\]

When the grating is projected on to a general surface with height \( h(x, y) \) the principal ray \( PCH \) strikes the surface at point \( H \), and point \( H \) will be seen at point \( D \) on plane \( R \) when observed through the imaging lens. The deformed grating \( g(x) = a(x, y) + b(x, y) \cos[2\pi f_0 x + \varphi(x)] \) encodes the height information in the phase-modulation term \( \varphi(x) = 2\pi f_0 x BD \). From the similarity of triangles \( \Delta HCD \) and \( \Delta HPI \) we obtain

\[
h(x, y) = \frac{l_0 \overline{CD}}{d_0 - CD}
\]

Noting that \( \overline{CD} = BD - BC \) we have \( \overline{CD} = (2\pi f_0)^{-1} \Delta \varphi(x) \). and substituting into equation 1-38 gives.
\[ h(x, y) = \frac{l_0 \Delta \phi(x)}{2 \pi f_0 d_0 - \Delta \phi(x)} \]  

(1-39)

where \( \Delta \phi(x) = \phi(x) - \phi_0(x) \) is the phase difference observed between the surface grating \( g(x, y) \) and the reference grating \( g_0(x, y) \).

Discontinuous surfaces may cause \( 2\pi \) ambiguity in the recovered phase unless the projected phase range across the surface is limited to a single fringe period, which severely limits the measurement accuracy. Practical systems overcome this problem by combining: (a) a phase shifting approach to determine the phase modulo \( 2\pi \); and (b) a hierarchical approach to determine the missing integral multiple of \( 2\pi \) that must be added to obtain the absolute phase (equivalent to the determination of the correct fringe order at each pixel). The general principle consists of the generation of a synthetic spatial wavelength \( \lambda \) along the \( x \)-axis by tuning one of the flexible system parameters, for example: Saldner et al.[84] describe tuning the angle of one of the mirrors in a Michelson interferometer; Valera et al.[85] describe tuning the separation of two coherent fibre optic sources; Zhang et al.[86] describe tuning the orientation of a fixed grating in its own plane illuminated by incoherent light. The synthetic wavelength is scanned during the measurement procedure, increasing in a step-wise manner from a single fringe across the field of view, to obtain several phase maps at different sensitivities and thus enables improved accuracy whilst maintaining the dynamic range. Storage and subsequent interpretation of the large number of images produced by the technique can, however, require expensive hardware and significant computational effort.

Incoherent or coherent (interference) sinusoidal fringes can be used. Incoherent fringes are usually more convenient for projection on large objects but are strongly limited by depth of field of the projection optics. High-brightness, high-contrast sequences of computer generated fringe patterns can be projected at high speed using low-cost liquid crystal (LC) and digital mirror device (DMD) spatial light modulators, enabling optimisation of the fringes based on feedback of the acquired image sequence with only a small time penalty[87]. Coherent fringes introduce additional speckle noise but do not suffer from finite depth of focus and hence can be projected
into an entire measurement volume. As with other coherent systems, however, coherent fringe projection is susceptible to noise vibration and air currents.

1.3.11 Multiplexed spatial frequency of projected structured light

Although discontinuous surfaces can be measured by binary coding and spatial frequency scanning of structured light patterns the time required to project and record multiple frames restricts their application to surfaces that are static or in relative slow motion. Takeda et al.[88] describe a technique based on projected structured light that requires only a single fringe pattern and permits phase unwrapping for discontinuous surfaces from a single acquired image. Rather than temporal scanning of the grating spatial frequency as described in section 1.3.10, multiple phase maps with various phase sensitivities are spatial-frequency multiplexed into a single fringe pattern to unambiguously determine the surface height. The acquired image is filtered in the two dimensional Fourier domain to demultiplex the spatial carrier frequencies, each of which has been phase modulated by the surface height with varying sensitivity. The phase is decoded using the Fourier transform technique but will be wrapped into the range \([-\pi, \pi]\). The unambiguous phase must then be determined from the simultaneous wrapped phase equations using a modified phase unwrapping method.

By virtue of its single-shot recording the technique is suitable for the instantaneous measurement of dynamic motion that would not be possible using sequential projection and recording of multiple structured light patterns. However, multiplexing of spatial frequencies in the projected light pattern reduces fringe contrast and introduces difficulties during phase extraction. In particular, filtering of the acquired image in the Fourier domain band-limits the broadband signal obtained for discontinuous surfaces, and therefore reduces the attainable signal to noise ratio compared to sequential projection and recording schemes.

1.4 Summary

This chapter has presented a brief description of the most popular whole-field optical techniques for measuring surface displacement and surface profile. However, these represent only a small fraction of the wide range of optical arrangements that are described in the literature.
Digital speckle pattern interferometry has been widely adopted as the standard technique for displacement measurement in many application areas. Its acceptance has been fuelled both by availability of low-cost high-resolution CCD cameras and ever increasing performance-to-price ratio of personal computers which enables rapid quantitative analysis of the resultant interferograms. The technique offers the advantage of being able to perform non-invasive measurement to the sub-micron level in real-time at video rates enabling continuous deformation of objects under time-varying loads to be analysed.

Techniques for surface profile measurement have been described offering various trade-offs between dynamic range and sensitivity. Close range photogrammetry enables measurement of large surface areas with the ability to self-calibrate and self-check; however, is unsuitable for real-time implementation due to the computational effort required. Contouring techniques based on interference provide very high accuracy over short measurement ranges. Contour speckle interferometry remains popular since the simple optical arrangement can also be used for surface displacement measurements. Low coherence interferometry utilises the short coherence length of white light illumination to determine the optical path length difference between interferometer arms and therefore does not suffer from shading problems making it suitable for surfaces with deep holes. However, contouring speckle interferometry and low-coherence interferometry both require expensive mechanical devices to accurately position the test surface relative to the optical arrangement which restrict the speed of data acquisition. Fourier transform speckle profilometry overcomes these limitations by use of a frequency tuneable laser diode that enables fast scanning of the optical wavelength, however, the height sensitivity is limited by the quasi-linear scanning range between laser mod-hops. A Fabry-Perot etalon is used to filter a broadband light source in the spectral interference microscope and therefore provides a wider scanning range that allows improved height sensitivity. Coherent measurement techniques are susceptible to air currents and noise vibration, thus making them difficult to deploy in industrial environments. The co-axial co-image plane projection and observation technique utilises incoherent grating projection to provide a more robust method for profile measurement that is free from shading problems. However, an accurate mechanical translation device is required to
position the surface. Profile measurement in the medium- to long-range is dominated by active triangulation systems. Binary coding of sequences of projected structured light enables low-cost, high-speed characterisation of discontinuous surface profile, however, the accuracy is restricted by a finite number of codewords. Spatial frequency scanning of projected gratings overcomes this problem using the continuous phase parameter to identify projection planes. Spatial-frequency multiplexing of the projected gratings enables profile measurement of dynamic surfaces at the cost of reduced signal to noise.

Ultimately, the optical technique chosen to measure the physical property under test will depend on the required sensitivity and dynamic range. There is increasing interest in the ability to extract quantitative data regarding this physical quantity from the interference patterns produced by the optical methods described in this chapter. Robust algorithms, which are both efficient and resistant to errors, are being actively developed to perform this analysis and are discussed in detail in chapter 2.
1.5 Figures

Figure 1-1: Optical arrangement for recording the initial object state on the holographic plate.

Figure 1-2: Optical arrangement for recording of the loaded object state on the holographic plate.
Whole-field optical metrology

Virtual images (both states)

Reconstructed wavefronts

Virtual images (both states)

Figure 1-3: Optical arrangement for reconstruction of a double exposure holographic interferogram

Test object (loaded state) and virtual image (initial state)

Reconstructed and reflected wavefronts

Figure 1-4: Optical arrangement for reconstruction of a real-time holographic interferogram
Whole-field optical metrology

Figure 1-5: Optical arrangement for out-of-plane speckle interferometry

Figure 1-6: Optical arrangement for in-plane speckle interferometry
Figure 1-7: Optical arrangement for speckle shearing interferometry

Figure 1-8: Central perspective projection
Whole-field optical metrology

Figure 1-9: Optical arrangement for speckle contouring: geometrical analysis

(a)

(b)

Figure 1-10: Optical arrangement for low-coherence interferometry
Figure 1-11: Optical arrangement for Fourier transform speckle profilometry

Figure 1-12: Optical arrangement for co-axial co-image plane projection and observation profilometry
Figure 1-13: Optical arrangement for binary coding of projected structured light from a spatial light modulator (SLM) with crossed projection and imaging axes. Diagram shows one of a sequence of structured light patterns used in a 3-bit Gray code scheme to label the projection planes.

Figure 1-14: Optical arrangement for projected synthetic grating from a spatial light modulator (SLM) with crossed projection and imaging axes. Diagram shows one of a sequence of structured light patterns with zero phase on the projection axis and a uniform phase distribution from $-3\pi$ to $+3\pi$ across the projection field.
Figure 1-15: Optical geometry for projected structured light methods with crossed optical axes.
2 Quantitative interferogram analysis

2.1 Introduction
The common objective of most optical metrology techniques is to produce a two-dimensional intensity pattern (or interferogram) phase-modulated by the physical quantity being measured. The observed fringes typically represent contours of constant displacement component or depth, depending on the optical arrangement used to acquire the images. Although visual analysis of the fringe pattern enables qualitative diagnostics to be performed there is an increasing desire for quantitative results. This chapter provides an overview of popular techniques used to automatically recover quantitative information describing the physical parameter from such interferograms.

A sinusoidal interferogram may be represented by the following continuous intensity function:

\[ I(x, y) = I_0(x, y) + I_M(x, y) \cos \phi(x, y) \]  

(2-1)

where \( I_0(x, y) \) is the background illumination, \( I_M(x, y) \) is the intensity modulation, \( \phi(x, y) \) is the phase term related to the physical quantity being measured, and \( (x, y) \) are the spatial co-ordinates in the reference frame of the image. In general, \( I_0 \) and \( I_M \) are slowly varying functions of the spatial co-ordinates determined by variations in the reflectivity of the test surface and non-uniformity of the illumination. However, in the case of speckle correlation fringes that represent the intensity difference of speckle fields, the interferogram appears very noisy with the \( I_0 \) and \( I_M \) varying rapidly across the image as random variables (see appendix A1). The interferogram is typically imaged onto a CCD camera consisting of an array of photodetectors that spatially sample the two-dimensional signal, which is then digitised by a digital to analogue converter and stored on a frame grabber for further analysis in a digital computer. Sampling of successive images results in a three-dimensional intensity distribution \( I(m, n, t) \), where \( m = 0, 1, 2, \ldots, N_m - 1 \), \( n = 0, 1, 2, \ldots, N_n - 1 \), and \( t = 0, 1, 2, \ldots, N_t - 1 \). Equation 2-1 can be rewritten in terms of these non-dimensional spatial and temporal co-ordinates in its most general forms as follows:
Quantitative interferogram analysis

\[ I(m,n,t) = I_0(m,n,t) + I_m(m,n,t) \cos[\phi(m,n,t)] \]  

(2-2)

The inverse problem to reconstruct the spatial phase variable \( \phi(m,n,t) \) is mathematically ill-posed due to: (a) variations in the unknown background illumination and modulation terms; (b) ambiguity in the sign of the phase term due to parity of the cosine function i.e. \( \cos(\phi) = \cos(-\phi) \); and (c) wrapping of the phase term modulo \( 2\pi \) since \( \cos(\phi) = \cos(\phi + 2\pi \nu) \) where \( \nu \in \mathbb{Z} \) (i.e. \( \nu \) is a member of the set of positive and negative integers). It is not possible, therefore, to uniquely determine the phase variable from a single intensity observation.

Digital image processing of interferograms became popular with the increasing performance-to-price ratio of personal computers in the 1980s. During the last two decades a large number of computer-aided techniques for solving the inverse problem have been described in the literature and can broadly divided into passive and active methods. Early attempts at computer-aided analysis of the interferograms (see references [89] and [90], for example) were based on passive analysis of two-dimensional intensity distributions using conventional image processing techniques. The fringe maxima and minima are identified by a process of skeletonisation or fringe tracking and the phase map reconstructed by interpolation. Skeletonisation typically involves either: (a) numerical erosion of the fringes; or (b) thresholding followed by calculation of the local centre of mass. Fringe tracking uses special algorithms designed to “follow” paths along the maximum and minimum intensity regions that define the fringe maxima and minima. These techniques, however, assume local uniformity of \( I_0 \) and \( I_m \) and are thus unsuitable for the noisy interferograms typically obtained from speckle interferometry. Regularisation methods[91] are required to solve ambiguity of the fringe orders by integrating a-priori information of the physical variable, such as spatial continuity and smoothness, in order to obtain a unique and viable solution.

Although passive intensity based techniques are still sometimes used, methods to extract the phase distribution, either of the fringes or of the speckles themselves, by actively modifying the phase are now more common. These methods add to the phase function a known phase ramp, or carrier, which is linear in either time or position to
provide additional information to solve the regularisation and sign ambiguity problem. The former method, so called *temporal phase shifting*, is historically the most popular and samples intensity distributions at discrete time intervals to produce a three-dimensional intensity distribution. In the latter case, so called *spatial phase shifting*, a single two-dimensional interferogram is obtained. These methods offer significant advantages over the intensity based fringe analysis: (a) full-field data is obtained, not just at the fringe maxima and minima; (b) the sign of the phase distribution is obtained unambiguously; (c) local uniformity of $I_0$ and $I_M$ are not assumed and are thus more suitable for the analysis of speckle interferograms; and (d) noise immunity is generally better. The phase shifting technique enables measurement of wavefront deviations in the $\lambda/100$ region and repeatability up to around $10^{-2}\lambda$ to be achieved. The phase estimation process is required to solve a set of simultaneous equations for the three unknowns $I_0$, $I_M$, $\phi$ and thus requires at least a minimum of three measured intensity values.

In general we are interested in the change in some physical parameter between an initial and final state. The next step in the analysis therefore involves computing the phase change relative to the initial reference interferogram that might correspond, for example, to the undeformed state of the test surface. In particular, this step is required when calculating phase directly from speckles since the initial phase of a given speckle is random. Phase estimate values $\hat{\phi}$ evaluated using the arctangent function will lie in the range $[-\pi, \pi]$ and therefore the phase change values $\Delta\hat{\phi}$ will lie in the range $[-2\pi, 2\pi]$. These values are usually wrapped back in the principle range $[-\pi, \pi]$ before proceeding with the phase unwrapping process described below.

Temporal and spatial phase shifting techniques for analysing fringe patterns result in the estimated wrapped phase-change values (modulo $2\pi$ of the true phase-change) in the range $[-\pi, \pi]$ due to the arctangent function used in the phase-estimation process. The relationship between the wrapped phase and the unwrapped phase may be stated as

$$\phi(m,n,t) = \phi_w(m,n,t) + 2\pi v(m,n,t)$$ (2-3)
where $\phi(m,n,t)$ is the unwrapped phase, $\phi_w(m,n,t)$ is the wrapped phase, and $\nu(m,n,t)$ is an integer valued correcting field. A process of phase unwrapping is therefore required to remove the $2\pi$-phase discontinuities by addition of the correct integral multiple of $2\pi$ to each phase value. This is equivalent to the problem of assigning fringe orders in intensity-based fringe analysis.

In order to unwrap a given wrapped phase distribution $\phi_w(m,n,t)$ correctly, the original phase function must have been sampled in accordance with the Shannon sampling theorem, i.e. at least two samples per cycle. If the requirement is satisfied then the true phase change between two neighbouring sample points lies in the range $[-\pi, \pi]$ and the unwrapping problem then becomes trivial for phase maps calculated from good quality interferograms: the absolute phase difference between neighbouring wrapped phase estimates along the unwrapping axis is less than $\pi$ except for those regions with $2\pi$ discontinuities enabling the corrective field $\nu(m,n,t)$ to be determined. Unwrapping becomes more difficult when the absolute phase difference between neighbouring wrapped phase estimates is greater than $\pi$ at points other than discontinuities in the arctangent function. These erroneous discontinuities may be introduced by, for example, intensity noise, discontinuous surfaces, localised under-sampling, and speckle decorrelation. A large family of algorithms have been developed in recent years that are able to correctly identify these erroneous discontinuities and perform the unwrapping process in adverse conditions with varying levels of success. The majority of algorithms can be classed as either a temporal phase unwrapping or spatial phase unwrapping algorithm according to whether the unwrapping procedure is performed along the time axis or one or more of the spatial axes, respectively. Historically, temporal phase unwrapping has not been widely implemented simply due to the large three-dimensional data sets produced, however, this viewpoint is now changing in part due to the continued increase in performance to price ratio of modern computers.

The final step in the analysis involves conversion of the unwrapped phase map to the physical quantity of interest and is dependent on the optical configuration used to record the interferograms. For displacement fields, for example, this often involves only a simple scaling of the result. In addition, a post-processing stage may be
Quantitative interferogram analysis

required to transform the data from the pixel co-ordinate system \((m,n)\) onto the test surface co-ordinate system taking into account optical magnification and lens distortion.

2.2 Temporal phase shifting

The temporal phase shifting technique\([92,93]\) typically involves introducing a controlled linear phase ramp \(\varphi(t)\) against time and recording a sequence of interferograms. In interferometry arrangements the active phase modulation is conveniently achieved using a mirror mounted on a piezoelectric translator (PZT) to reflect either the object- or reference-beam. Alternatively, PZT driven glass wedges placed in the beam path can be used and are less likely to introduce lateral translation or tilt of the beam\([94]\). Optical fibres wrapped around PZT cylinders can also be used in fibre-based interferometers\([95,96]\). Although PZT devices are attractive due to their low-cost and low-voltage requirements these mechanical components are unsuitable for applications that demand high sampling rates. Pockels cells with response times in the order of nanoseconds have recently been demonstrated for dynamic ESPI problems\([97]\). Frequency tuneable diode laser based interferometers are also attractive for high sampling rate applications since phase shifting can be achieved without the requirement for further physical devices if the interferometer arms are arranged with unequal path lengths\([98]\).

Temporal phase shifting estimates the phase at each pixel in the interferogram independently of all other pixels and assumes that changes in \(I_o\) and \(I_M\) (for example, due to speckle decorrelation), and \(\phi\) (for example, due to surface motion) are negligible within a small number of frame periods \((\tau)\). The technique is therefore susceptible to external disturbances like vibration or rapid changes in temperature and is unsuitable for measurements of dynamic events. In speckle interferometry we can assume the complex amplitudes of both interfering fields remain constant within the finite area of the pixel if the detector element is sufficiently small to resolve the speckle. In this case each pixel works as a classical interferometer with a background intensity and modulation of its own. A theoretical study of temporal phase shifting for
Quantitative interferogram analysis

an arbitrary number of speckles per pixel is provided by Lehmann\cite{99} and demonstrates that good phase estimates can also be obtained for unresolved speckles.

Incorporating the linear phase ramp in equation 2-2 and dropping the spatial dependence gives

\[ I(t) = I_0(t) + I_M(t) \cos[\phi(t) + \psi(t)] \] (2-4)

where \( \psi(t) = \psi t \), and \( \psi \) represents the phase shift introduced between two successive frames. The phase ramp introduces a temporal carrier frequency \( \nu_0 = \psi / (2\pi t) \) that enables the sign of changes in the physical quantity to be determined from positive or negative variations in the signal frequency relative to the carrier frequency \( \nu_0 \).

More precisely, the term phase shifting is used to denote those cases in which the phase ramp is continuous throughout the camera exposure period, and phase stepping used in those cases where the phase is increased in discrete steps between frames but held constant during the exposure period. The result of integrating the intensity over a continuous phase ramp in the phase shifting approach can be calculated as

\[ I(t) = \frac{1}{\tau} \int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} I_0 + I_M \cos[\phi + \psi(t)] \, dt \] (2-5)

If we assume a linear phase ramp \( \psi(t) = \psi t \) then the integral can be evaluated by substitution, and using the formula \( \sin(\alpha + \beta) - \sin(\alpha - \beta) = 2\cos\alpha\sin\beta \) gives

\[ I(t) = I_0 + \text{sinc} \left( \frac{\Delta\phi}{2\pi} \right) I_M \cos[\phi + \psi(t)] \] (2-6)

where \( \Delta\phi = \psi \tau / 2 \) is the phase shift introduced during one frame period.\textsuperscript{1} It can be seen therefore that the phase modulation in phase shifting is somewhat reduced when compared to phase stepping due to the sinc function, however, subsequent processing steps are identical for each approach\cite{100}.

The following sections introduce a number of popular schemes for estimating the phase from a small number of intensity samples \( M \) taken over just one or two cycles of the carrier wave. The precision of the phase estimate is affected by a number

\[ \text{sinc}(x) = \frac{\sin(x)}{x} \]
of factors including mis-calibration of the phase modulator, non-linearity effects in the phase modulator and photodetector, and environmental disturbances. The choice of phase stepping algorithm is therefore a compromise between sensitivity to these various error sources and computational effort.

### 2.2.1 Four-frame algorithm

The most popular phase estimation scheme is obtained by considering the case where $M = 4$ and $\psi = \pi / 2$ and leads to the simplest implementation. Substituting these values and $t = 0, 1, 2, \text{and} 3$ into equation 2-4 and dropping the spatial dependence results in the following set of equations:

\[
\begin{align*}
I(0) &= Io + IM \cos \phi \\
I(1) &= Io + IM \cos(\phi + \pi / 2) = Io - IM \sin \phi \\
I(2) &= Io + IM \cos(\phi + \pi) = Io - IM \cos \phi \\
I(3) &= Io + IM \cos(\phi + 3\pi / 2) = Io + IM \sin \phi
\end{align*}
\]

(2-7)

Estimates for both phase $\phi$ and modulation $I_M$ at time zero can be calculated by rearranging equation 2-7 as follows

\[
\phi(0) = \tan^{-1} \left( \frac{I(3) - I(1)}{I(0) - I(2)} \right) \quad \text{(2-8)}
\]

\[
i_M(0) = \sqrt{[I(3) - I(1)]^2 + [I(0) - I(2)]^2} / 2 \quad \text{(2-9)}
\]

where $\hat{\ }$ is used to denote that these terms are estimators for the true value and may differ due to random and systematic measurement errors. The four-frame algorithm is attractive due to its computational simplicity, however, in practice is susceptible to errors due to mis-calibration of the phase modulator. A linear mis-calibration such that the actual phase ramp gradient is $\psi' = \psi(1 + \varepsilon)$ results in a phase estimation error given by

\[
\delta \phi = \hat{\phi} - \phi = \pi \varepsilon (3 + \cos 2\phi) / 4 \quad \text{(2-10)}
\]

to the first order in $\varepsilon$ [101]. The constant term will cancel when calculating phase changes, however, in general the cosine term will not and results in a phase error that is proportional to the linear mis-calibration.
2.2.2 Five-frame algorithm

One approach used to reduce phase errors due to mis-calibration of the phase modulator is the so-called averaging technique proposed by Schwider et al[102]. A new error-compensating algorithm can be calculated from an existing algorithm if two sets of data with a $\pi/2$ phase shift in the initial phase are taken. In general, this process may require the acquisition of twice as many frames, however, in the case of algorithms based on $\psi = \pi/2$ phase shift between frames then only one extra frame is required since the first data set will overlap with the second. The five-frame algorithm developed from the four-frame algorithm described in section 2.2.1 can be written as follows:

$$
\hat{\phi}(0) = \tan^{-1}\left( \frac{2[I(3) - I(1)]}{I(4) + I(0) - 2I(2)} \right)
$$

$$
\hat{\psi}_M(0) = \sqrt{4[I(3) - I(1)]^2 + [I(4) + I(0) - 2I(2)]^2} / 4
$$

The phase error due to mis-calibration is now given by[101]

$$
\delta\phi = \hat{\phi} - \phi = \pi\varepsilon + (\pi\varepsilon / 4) \sin 2\phi
$$

where the twice-the-fringe-frequency phase term is now seen to scale quadratically with $\varepsilon$, rather than linearly as with the four-frame algorithm, and is therefore significantly attenuated.

2.2.3 Carré technique

An alternative approach to the problem of mis-calibration of the phase modulator is the so-called self-calibrating technique in which the phase shift $\psi$ is regarded as one more unknown variable that must be estimated. A minimum of four intensity samples is therefore required to solve the four unknowns. The classical Carré solution is based on four samples and assumes that the unknown phase step between samples is constant (i.e. linear). Using a carrier phase signal $\psi(t - 3/2)$ gives phase shift values of $[-3\psi/2, -\psi/2, \psi/2, 3\psi/2]$ and the estimated phase shift can be written as:

$$
\psi = 2\tan^{-1}\left[ \frac{3[I(1) - I(2)] - [I(0) - I(3)]}{[I(1) - I(2)] + [I(1) - I(4)]} \right]
$$

Once the value of $\psi$ has been calculated, the signal phase can be estimated as:
Quantitative interferogram analysis

\[
\phi(0) + \frac{3\psi}{2} = \phi(3/2) = \tan^{-1}\left\{ \tan\left(\frac{\psi}{2}\right) \left[ \frac{[I(0) - I(3)] + [I(1) - I(2)]}{[I(1) + I(2)] - [I(0) + I(3)]} \right] \right\} \tag{2-15}
\]

where the term \(3\psi/2\) on the left hand side is due to the choice of the carrier origin and cancels when calculating phase difference values. Substituting equation 2-14 into equation 2-15 gives

\[
\hat{\phi}(3/2) = \tan^{-1}\left\{ \frac{\text{sign}[I(1) - I(2)]\sqrt{[I(0) - I(3)] + [I(1) - I(2)]} [3[I(1) - I(2)] - [I(0) - I(3)]}}{[I(1) + I(2)] - [I(0) + I(3)]} \right\}
\tag{2-16}
\]

where \(\text{sign}(x) = 1\) for \(x \geq 0\), and \(-1\) otherwise. Whilst this technique provides a practical solution for mis-calibration errors the resulting algorithm is mathematically complex, requiring the computation of square roots in addition to the usual arctangent function, and becomes unstable as \(\psi\) approaches \(\pi\) since both numerator and denominator drop to zero. Furthermore, the time of the computed phase estimate does not correspond to that of any of the four intensity samples.

### 2.2.4 Fourier-transform representation

A recent trend in phase shifting algorithm design follows from the Fourier description of the algorithm[103]. In this case the algorithm is viewed as a filtering process in the frequency domain allowing various algorithms to be compared according to their frequency response. The frequency-domain approach is attractive since it provides a more general approach to algorithm development than the previous methods.

Consider the hypothetical case of a periodic continuous intensity signal given by

\[
I(t) = I_0 + I_{sl} \cos[\phi(t) + \varphi(t)]
\tag{2-17}
\]

where \(\phi(t)\) is the interference phase to be estimated, and \(\varphi(t)\) is the phase step. It is assumed that \(\phi(t)\) is a slowly varying function that can be estimated by a constant \(\hat{\phi}\) over one cycle of the carrier. The Fourier transform is given by

\[
\tilde{I}(\omega) = \int_I I(t) W(t) \exp(-i2\pi\omega t) \, dt
\tag{2-18}
\]

where \(\omega\) is the continuous temporal frequency variable and \(W(t)\) is a continuous window function. Assuming that the phase shift is continuous and linear such that
Quantitative interferogram analysis

\[ \varphi(t) = 2\pi \omega_0 t \]  

(i.e. \( \psi = 2\pi \omega_0 \)) then the Fourier transform equates to

\[
\tilde{I}(\omega) = I_0 \tilde{W}(\omega) + \frac{1}{2} \left[ \tilde{W}(\omega - \omega_0) \exp(i\phi) + \tilde{W}(\omega + \omega_0) \exp(-i\phi) \right]
\]

(2-20)

where \( \tilde{W}(\omega) \) is the Fourier transform of the window function[104]. A schematic representation of equation 2-20 is shown in figure 2-3(a) where \( \tilde{I}(\omega) \) is evaluated as the weighted sum of three Fourier transforms of the window function. The weighting of each transform is represented by the three delta functions centred at \( \omega = -\omega_0, 0, \) and \( \omega_0 \). If we consider the case of a rectangular window function \( W(t) \) of duration equal to one complete period of the carrier then the Fourier transform \( \tilde{W}(\omega) = \text{sinc}(\omega) \).† Figure 2-3(a) includes only the sinc function centred on \( \omega = -\omega_0 \) for clarity. In the absence of spectral leakage (i.e. \( \tilde{W}(\omega_0) = 0 \) and \( \tilde{W}(2\omega_0) = 0 \)), the interference phase can thus be estimated as the argument of the Fourier coefficient at \( \omega = \omega_0 \) as follows:

\[
\hat{\phi} = \tan^{-1} \left\{ \frac{\text{Im} \left[ \tilde{I}(\omega_0) \right]}{\text{Re} \left[ \tilde{I}(\omega_0) \right]} \right\} + \text{const}
\]

(2-21)

The constant is the complex phase of \( \tilde{I}(0) \) and can normally be ignored since we are usually interested in phase-change values.

In practice phase shifting involves discrete sampling of the intensity signal. The discrete Fourier transform of the set of \( M \) sampled intensity values \( I(t) \) \((t=0, 1, 2, ..., M-1)\) is given by

\[
\tilde{I}(k) = \sum_{t=0}^{M-1} I(t) W(t) \exp(-i2\pi k t / N)
\]

(2-22)

where \( k \) is a continuous non-dimensional temporal frequency, and \( N \) is the number of samples per cycle of the carrier, i.e. \( N = 2\pi / \psi \). Equation 2-22 can be rewritten using the Dirac delta function as follows:

† The function \( \text{sinc}(x) \) is defined as \( \frac{\sin(\pi x)}{\pi x} \)
Quantitative interferogram analysis

\[ \tilde{I}(k_i) = \int_{-\infty}^{\infty} I(t)W(t) \left[ \sum_{i=0}^{\infty} \delta(t-i') \right] \exp(-i2\pi k_i t / N) \, dt \] (2-23)

where \( t \) is now a continuous variable. Using the convolution theorem we have:

\[ \tilde{I}(k_i) = \tilde{W}(k_i) \ast \sum_{j=-\infty}^{\infty} I_0 \delta(k_i - jN) + \frac{I_M}{2} \exp(i\phi) \delta(k_i - jN - 1) + \frac{I_M}{2} \exp(-i\phi) \delta(k_i - jN + 1) \] (2-24)

where \( \ast \) denotes convolution. A schematic representation of equation 2-24 is shown in figure 2-3(b). If we consider again the rectangular window function \( W(t) \) of duration equal to one complete period of the carrier then \( \tilde{W}(k_i) = \text{sinc}(k_i) \). The three terms within the summation corresponding to \( j = 0 \) occur at \( k_i = -1, 0, \) and \( +1; \) these peaks are aliased along the frequency axis at integral multiples of \( N \) (corresponding to the non zero values of \( j \)) due to the sampling process. Convolution with the transform of the window function is equivalent to placing a copy of \( \tilde{W}(k_i) \) centred on each of the delta functions and computing the sum. Figure 2-3(b) shows only the \( \tilde{W}(k_i) \) function centred on \( k_i = -1 \) for clarity. The interference phase can now be estimated as the argument of the Fourier coefficient at \( k_i = 1 \) as follows:

\[ \hat{\phi} = \tan^{-1} \left( \frac{\text{Im}[\tilde{I}(1)]}{\text{Re}[\tilde{I}(1)]} \right) \] (2-25)

Substituting \( \tilde{I}(1) \) from equation 2-22 into equation 2-25 gives the general phase shifting equation[92]:

\[ \hat{\phi} = \tan^{-1} \left( \frac{N(t)}{D(t)} \right) \] (2-26)

where:

\[ N(t) = \sum_{r=0}^{M-1} I(t)b(t) \] (2-27)

\[ D(t) = \sum_{r=0}^{M-1} I(t)a(t) \] (2-28)

and where the sampling coefficients are given by

\[ a(t) = \text{Re}[W(t) \exp(-i2\pi t / N)] \] (2-29)
Quantitative interferogram analysis

\[ b(t) = \text{Im}[W(t) \exp(-i2\pi t / N)] \] (2-30)

This approach enables phase shifting algorithms to be fully specified simply by the set of sampling coefficients \( \{a(t), b(t)\} \) and the number of samples per cycle \( N \). The sampling coefficients for the four-frame algorithm discussed in section 2.2.1, for example, can be calculated using \( M = N = 4 \) together with the rectangular window function \([W(0),\ldots,W(3)] = [1,1,1,1]\). Coefficients for algorithms applied in the following chapters are included in Table 2-1.

2.2.5 Errors in the phase estimator

The selection of optimum values for the sampling coefficients in equation 2-26 is based on minimising the systematic and random errors in the estimated phase for a given number of intensity samples. The principal sources of systematic error include: (a) higher harmonics present in the recorded intensity signal; (b) mis-calibration of the phase modulator; and (c) vibration.

In practical measurements the observed intensity signal may often depart from a pure sinusoid as assumed in section 2.2.4. Higher harmonics may be introduced from several sources, including non-linearities in the photodetector and multiple-beam interference due to reflections. Harmonics with a frequency greater than the Nyquist limit \( k_c = N/2 \) will be aliased onto the principal range \([-N/2,N/2]\) and may then interfere with the frequency at which the phase is evaluated, namely \( k_1 = 1 \). Harmonic components with a frequency \( q \) (where \( q \in \mathbb{Z} \)) times the fundamental will produce corresponding spectral peaks at \( k_i = jN \pm q \) (where \( j \in \mathbb{Z} \)) and therefore significant phase errors can be expected if \( N = q \pm 1 \). Hence, the phase shifting algorithm should be chosen such that \( N > q_m + 1 \) where \( q_m \) is the maximum frequency component in the original intensity signal[18,105].

Mis-calibration of the phase modulator results in a departure of the carrier frequency from the expected value \( \omega_0 \). A fractional linear phase-shift error \( \varepsilon \) causes the fundamental peaks \( S_0^+ \) and \( S_0^- \) to shift sideways from \( k_c = \pm 1 \) and appear at \( k_i = \pm(1 - \varepsilon) \) instead. The zero of the sinc function centred on the peak \( S_0^- \) no longer coincides with the point at which the phase is evaluated \( k_i = 1 \) and spectral leakage
from the neighbouring peak thus causes systematic errors in the phase estimate. These phase errors can be reduced by choosing an improved window function in the time domain to reduce the effect of spectral leakage in the frequency domain. The $M = N + 1$ class of algorithm[106] (of which the five frame algorithm described in section 2.2.2 is an example) is particularly efficient since it requires only one extra sample, however, increasing the number of samples to a minimum of $M = 2N - 2$ is recommended in the presence of significant harmonic content[107].

A generally acknowledged drawback of phase shifting is its sensitivity to vibration and other environmental disturbances. de Groot[108,109] has developed numerical methods to calculate the frequency response of the root mean square (RMS) phase error for a given phase shifting algorithm, thus enabling algorithms to be selected that are insensitive to the vibration frequencies observed. The largest phase error is produced at a vibration frequency equal to twice the fundamental carrier frequency $\omega_0$ and most common phase-shifting algorithms are equally sensitive to vibration at this resonant frequency. It is often possible to perform accurate phase evaluation without recourse to mechanical isolation provided that the vibration amplitude is small and its spectral content can be characterised. In general, increasing the sampling rate as high as possible with respect to the vibration frequency reduces the effects of vibration. Huntley et al.[97] have demonstrated a phase-shifted ESPI system with a 1 kHz framing rate, however, high-speed phase modulators and photodetectors with short integration times are relatively expensive. Alternative hardware (such as feedback loops) and software techniques (reference [110], for example, calculates a corrective term derived from spectral analysis of the time-varying intensity signal) have been proposed to reduce errors in the estimated phase due to vibration. More conventional approaches include the use of quasi-common-path interferometers and systems based on non-coherent signal analysis.

Random errors are also present in practical measurements, and may be introduced by electronic noise in the photodetector, random variations in the illumination intensity, and poor signal modulation. The latter source is particularly problematic in speckle interferometry arrangements with unresolved speckles. Surrel[111] showed that when searching for an optimum phase shifting algorithm in the presence of random intensity
errors, error-resilience is dominated by the number of intensity samples, with the average standard deviation in estimated phase scaling as $1/\sqrt{M}$, and that the choice of sampling coefficients has only limited effect.

### 2.2.6 Takeda Fourier transform method

The Fourier transform technique for fringe pattern analysis, first proposed by Takeda et al. [112] in 1982, has been widely applied to the analysis of interferograms with spatial carriers, and more recently, to phase estimation in speckle interferograms with temporal carriers [113,114]. The interference phase is estimated using the following procedure: (a) the sampled phase modulated intensity signal is transformed into the Fourier domain; (b) the frequency spectrum is asymmetrically bandpass filtered; (c) the inverse transform is taken; and (d) the real and imaginary parts of the resulting signal are compared.

In practice the interference phase $\phi(t)$ is estimated from $N_i$ samples of the discrete intensity signal $I(t)$ using the discrete Fourier transform, defined as:

$$\tilde{I}_i(k_i) = \sum_{i=0}^{N_i-1} I(t) \exp(-i2\pi k_i t / N_i) \quad (2-31)$$

If we consider an intensity signal modulated by a linear phase ramp $\varphi(t) = \psi t$, then following the arguments discussed in section 2.2.4, the Fourier domain spectrum will consist of peaks centred at $k_i = -jN_i$ and $k_i = \pm k_0 - jN_i$, where $k_0 = \psi N_i / 2\pi$ and $j \in \mathbb{Z}$. Spectral leakage between the peaks will be prevented if variation in the interference phase over $N_i$ samples is small, and that the carrier frequency is sufficiently high. The transform is multiplied by a window function $\tilde{W}_i(k_i - k_0)$ chosen to isolate the first positive-frequency side-lobe, which is then shifted along the $k_i$ axis by $-k_0$ to demodulate the signal. The inverse Fourier transform is applied to obtain

$$I_F(t) = \frac{I_M(t)}{2} \exp[i\phi(t)] \quad (2-32)$$

and the phase estimate evaluated as follows:
This frequency domain evaluation approach for estimating phase provides very good agreement with those results obtained using time domain analysis provided that \( N_r = M \) and equivalent window functions are chosen. This similarity is derived from the fact that the numerical procedures are mathematically analogous.

### 2.3 Spatial phase shifting

The temporal phase shifting method described in the preceding sections requires the recording of at least three phase-modulated interferograms that are recorded sequentially. Disturbances introduced by thermal and mechanical fluctuations during the recording time introduce errors in the estimated phase. Spatial phase shifting provides an alternative approach that can be regarded as simultaneous multi-channel recording of a phase-modulated signal by use of either several cameras with an appropriate static phase shift for each of the images or by the introduction of a carrier fringe pattern on a single camera target. Such multi-channel recording techniques eliminate errors introduced by temporal disturbances and are intrinsically suitable for measurement of high-speed events due to their single-frame measuring ability. Errors can be introduced, however, by variations in the sensitivity of the different photodetectors used to measure the phase-modulated intensity field. Subsequent analysis of the sampled phase-modulated signal is identical to that applied in temporal phase shifting.

#### 2.3.1 Spatially-separated interferograms

One method successfully applied for simultaneous recording the phase-modulated signal involves spatially separating each of the phase-shifted interferograms. Diffraction gratings or beam splitters are used to generate three or four phase-shifted interferograms on separate regions of the photodetector array, or even separate cameras[115,116]. In general, such systems are very expensive and require the use of a high power light source due to wavefront-division of the object beam. The system is impractical for most speckle interferometry applications, however. Since sub-pixel
alignment of the targets over the entire field is required to prevent speckle decorrelation.

2.3.2 Speckles with spatial carrier

A second approach to spatial phase shifting involves recording only a single image in which a spatial carrier has been introduced. In an out-of-plane interferometer, for example, the carrier can be introduced by oblique translation of the origin of the reference wave away from the centre of the imaging aperture before it is superposed upon the object wave thus generating a linear spatial phase-shift between the two wavefronts[117]. If we consider a two-dimensional intensity field

\[ I(m, n) = I_0(m, n) + I_M(m, n) \cos[\phi(m, n) + \varphi(m, n)] \] (2-34)

then a translation of the reference wavefront in the horizontal direction will generate a phase modulation function \( \varphi(m, n) = 2\pi f_0 m \), where \( f_0 \) is the spatial frequency of the resulting vertical carrier fringes. The local interference phase can then be estimated from any set of three horizontally adjacent pixels across the interferogram (except the edges) provided that local spatial variations in \( I_0 \) and \( I_M \) are negligible, and that the intensity signal is sampled in accordance with the Shannon sampling theorem, i.e. at least two samples per cycle.

The equivalence of equations 2-4 and 2-34 implies that all of the phase shifting algorithms applied in the domain of temporal phase shifting are equally applicable for the analysis of interferograms with spatial carriers. In practice, a \( M \)-sample algorithm is implemented by separately convolving the interferogram with two \( M \times 1 \) kernels consisting of the \( a(t) \) and \( b(t) \) sampling coefficients of equations 2-29 and 2-30. The phase estimate can then be computed as the arctangent of the ratio of the two convolved images. The convolution operation results in the spatial resolution along one axis being degraded by a factor \( M \), and sensor arrays with higher resolution than those used for temporal phase shifting are required.

The application of spatial phase shifting in classical interferometry is relatively straightforward. In the presence of speckle, however, it is unlikely that the assumption of constant \( I_0, I_M \), and \( \phi \) over the local region of \( M \) sample points that is required by most phase shifting algorithms will be satisfied unless care is taken. In temporal
phase shifting, where the phase is calculated from samples along the time axis, this assumption is usually valid provided that temporal disturbances can be avoided and speckle decorrelation is controlled. By contrast, in spatial phase shifting where the samples are taken along one of the spatial axes, intensity and phase gradients within and between the speckles can introduce severe errors in the estimated phase. In general, errors due to these gradients will increase as the mean speckle size is reduced, however, large speckles will increase speckle decorrelation and provide poor light efficiency. As discussed in section 2.2, temporal phase shifting can be implemented successfully using unresolved speckles (i.e. several speckles per photodetector), whereas spatial phase shifting requires a mean speckle size covering at least $M$ photodetectors and therefore implies a reduced aperture that must be compensated for by an increase in laser power.

Burke et al.\cite{118} suggested that phase estimation errors due to phase gradients in interferogram are equivalent to linear mis-calibration of the phase modulator in temporal phase shifting (discussed in section 2.2.5) that has received much attention. The use of algorithms designed to be insensitive to linear mis-calibration of the phase modulator should therefore be chosen to reduce these errors. Furthermore, taking an additional recording of the object speckle field intensity (e.g. by blocking the reference wavefront) and incorporating these values into the interferogram equation enables the assumption of constant $I_0$ and $I_M$ to be relaxed. The combination of these methods allows relatively high precision measurements to be achieved in practice.

A recent paper by Burke et al.\cite{119} provides quantitative experimental data for comparing the performance of temporal and spatial phase shifting in ESPI. Various experimental parameters where investigated including object-illumination intensity, speckle size, and fringe density. The comparisons were made using a reference-to-object-beam intensity ratio of 10:1 as this was found to be optimal for both methods. In general spatial phase shifting was found to be inferior to temporal phase shifting, however, the relative performance was found to depend on the degree of speckle decorrelation and the spatial resolution of the photodetector array. In particular, the
performance advantage of temporal phase shifting decreases rapidly with increasing speckle decorrelation.

### 2.3.3 Fringe patterns with spatial carrier

Spatial phase shifting can also be performed using correlation fringe patterns. In this case the spatial carrier is introduced into the subtraction (or addition) fringe pattern by generating a linear phase variation across either of the two speckle interferograms. This approach is different from that described in section 2.3.2 in which it is actually the speckles that are phase modulated rather than the fringe pattern. In an out-of-plane interferometer, for example, phase modulation can be achieved by introducing a small tilt of the object beam between the two exposures. The technique is particularly suitable for measuring dynamic events using pulsed lasers[120-122]. The subtraction (or addition) fringe pattern can be represented as

\[
I(m, n) = I_0(m, n) + I_M(m, n) \cos[\Delta \phi(m, n) + \phi(m, n)]
\]

which differs from equation 2-34 in that it encodes \( \Delta \phi \) rather than \( \phi \). Subsequent processing step are almost identical, however, it should be noted that the estimated phase will still contain the spatial carrier. This is normally eliminated by subtraction of a phase map obtained when there has been no change in the physical parameter between exposures.

### 2.4 Temporal phase unwrapping

Phase estimation methods based on phase shifting generate values in the principal range \([ -\pi, \pi ]\) due to the arctangent function used during evaluation. Phase unwrapping is required to remove the \( 2\pi \) phase discontinuities present in the estimated phase signal by the addition of the correct integer multiple of \( 2\pi \) to each phase value. Temporal phase unwrapping calculates the integer valued correcting field \( v(m, n, t) \) by comparing successive phase estimates along the time axis. Temporal phase unwrapping is attractive as the analysis is performed using only a single dimension (i.e. time) and is therefore more straightforward than alternative methods based on comparing phase values along two or more spatial axes. However, all temporal phase unwrapping algorithms are based on the assumption that the original phase signal has been sampled along the time axis in accordance with the
Shannon sampling theorem and their use is thus restricted to those applications in which loading can be precisely controlled (i.e. the magnitude of the phase change should not exceed $\pi$ for each load increment). One consequence of this is that extended measurement sequences require a large number of loading steps and thus produce large data sets. The main advantage gained by unwrapping along the time axis is that erroneous phase values due, for example, to poor modulation, do not propagate spatially in the data set. Furthermore, physical discontinuities are automatically respected and isolated regions in the interferogram are correctly unwrapped without any uncertainty concerning their relative phase order. For example, figure 2-1(a) shows the wrapped phase map obtained using projected fringes to measure the surface profile of two physically separated turbine blades. Temporal phase unwrapping enables their relative phase origins to be determined unambiguously, as shown in figure 2-1(b), even though there is no noise-free spatial unwrapping path available. Unwrapping errors can propagate along the time axis, however, and the presence of a single corrupted phase map in the data set can prevent successful calculation of the final unwrapped phase. Sections 2.4.1-2.4.3 describe the three main algorithms used to implement temporal phase unwrapping. For the following discussion we define the function $\Delta \hat{\phi}_w(t_1, t_2)$ to represent the difference between wrapped phase values at times $t=t_1$ and $t=t_2$ rewrapped back into the principal range $[-\pi, \pi]$. This function can be written as

$$\Delta \hat{\phi}_w(t_1, t_2) = W[\hat{\phi}_w(t_1) - \hat{\phi}_w(t_2)]$$ (2-36)

where the wrapping operator, $W$, can be expressed as

$$W(\phi) = \phi - 2\pi \text{NINT}(\phi / 2\pi)$$ (2-37)

and NINT denotes rounding to the nearest integer.

### 2.4.1 Basic method

The simplest implementation of temporal phase unwrapping involves calculating the number of $2\pi$ phase jumps between two successive wrapped phase values:

$$d(t) = \text{NINT}\left[\hat{\phi}_w(t) - \hat{\phi}_w(t-1)\right] / 2\pi, \quad t=1, 2, \ldots, N_t-1$$ (2-38)

where $N_t$ is number of phase values along the time axis. Normally, the term $\hat{\phi}_w(t) - \hat{\phi}_w(t-1)$ will lie in the range $[-\pi, \pi]$, however, when a phase discontinuity
Quantitative interferogram analysis

is crossed, it will lie in the range \([-2\pi, -\pi]\) or \([\pi, 2\pi]\) and thus produce non-zero values of \(d(t)\). The total number of \(2\pi\) discontinuities up to the \(t^{th}\) phase value is thus obtained as

\[
\nu(t) = \sum_{t'=1}^{t} d(t'), \quad t=1, 2, ..., N, -1
\]

(2-39)

and the estimated total unwrapped phase change between the initial phase value and the \(t^{th}\) phase value, \(\Delta \hat{\phi}(t,0) = \hat{\phi}(t) - \hat{\phi}(0)\), is given by

\[
\Delta \hat{\phi}(t,0) = \hat{\phi}_w(t) - \hat{\phi}_w(0) - 2\pi \nu(t)
\]

(2-40)

Figure 2-2 illustrates addition of integer multiples of \(2\pi\) phase steps to obtain the unwrapped phase value. Whilst this simple approach is sufficient for analysing good quality wrapped phase maps, it is not easily adapted for use on data generated by speckle interferometers. Random changes in speckle amplitude produce spatial variations in the intensity modulation, resulting in very poor signal-to-noise at some pixels in the interferogram. Spatially smoothing of the phase difference map using a local average operation is often required, however, this cannot easily be incorporated into the current algorithm due to the presence of \(2\pi\) discontinuities in the term \(\hat{\phi}_w(t) - \hat{\phi}_w(t-1)\) from equation 2-38. Simple low pass filtering cannot be applied, for example, since these sharp discontinuities must be preserved for successful unwrapping to be achieved.

2.4.2 Sum of differences method

A second implementation was initially described by Itoh[123] as the cascading of the three simple operators: differentiation, wrapping, and integration; and separately derived by Huntley et al.[3]. Analysis is based on the property that the change over time of any function can be written as a sum of differences. When applied to the function \(\hat{\phi}(t)\), this gives

\[
\Delta \hat{\phi}(t,0) = \sum_{t'=1}^{t} [\hat{\phi}(t') - \hat{\phi}(t' - 1)]
\]

(2-41)

The term inside the square brackets will always lie in the range \([-\pi, \pi]\) provided that the phase function has been sampled in accordance with the Shannon sampling theorem (as assumed above). Consequently, the difference between successive
unwrapped phase values is equal to the rewrapped difference between successive wrapped phase values. This can be expressed as
\[
\hat{\phi}(t) - \hat{\phi}(t-1) = W[\hat{\phi}_w(t) - \hat{\phi}_w(t-1)] = \Delta \hat{\phi}_w(t, t-1)
\] (2-42)

Substituting equation 2-42 into equation 2-41 gives
\[
\Delta \hat{\phi}(t, 0) = \sum_{t'=1}^{t} \Delta \hat{\phi}_w(t', t'-1) \tag{2-43}
\]

In the absence of spatial smoothing, equation 2-43 will give identical values to those produced by the basic algorithm given in equation 2-40 (apart from numerical rounding errors that can normally be neglected). The intermediate unwrapped phase values of \( \hat{\phi}(t') \) (for \( t'=1, 2, \ldots, t-1 \)) in equation 2-41 cancel out when the summation is evaluated and so that errors in the intermediate estimates do not contribute to the final unwrapped phase value.

Burke et al. [124] noted that the composite function \( \Delta \hat{\phi}_w(t_1, t_2) \) can be calculated in a single step for all phase shifting algorithms. If the wrapped phase estimates are interpreted as complex numbers \( z(t) \) with modulus 1 and argument \( \hat{\phi}_w(t) \) such that
\[
z(t) = \cos \hat{\phi}_w(t) + i \sin \hat{\phi}_w(t) \tag{2-44}
\]
and
\[
\hat{\phi}_w(t) = \arg\{z(t)\} = \tan^{-1} \left( \frac{N(t)}{D(t)} \right) \tag{2-45}
\]
where
\[
N(t) = \text{Im}[z(t)], \quad D(t) = \text{Re}[z(t)] \tag{2-46}
\]
The difference between wrapped phase estimates can now be expressed as
\[
\arg\{z(t_1)\} - \arg\{z(t_2)\} = \arg\left( \frac{z(t_1)}{z(t_2)} \right) \tag{2-47}
\]
The important property of equation 2-47 is that the right hand side will automatically wrap into the range \([-\pi, \pi]\), which is equivalent to calculating \( \Delta \hat{\phi}_w(t_1, t_2) \) directly.

The complex division can be evaluated as
\[
\frac{z(t_1)}{z(t_2)} = \frac{N(t_1)N(t_2) + D(t_1)D(t_2)}{D(t_2)^2 + N(t_2)^2} + i \frac{N(t_1)D(t_2) + D(t_1)N(t_2)}{D(t_2)^2 - N(t_2)^2} \tag{2-48}
\]
and substituting equations 2-45 and 2-48 into equation 2-47 gives the general formula:

\[ \Delta \hat{\phi}_w(t_1, t_2) = \tan^{-1}\left\{ \frac{F(t_1, t_2)}{G(t_1, t_2)} \right\} \]  

(2-49)

where

\[ F(t_1, t_2) = N(t_1)D(t_2) - D(t_1)N(t_2) \]  

(2-50)

\[ G(t_1, t_2) = N(t_1)N(t_2) + D(t_1)D(t_2) \]  

(2-51)

The direct evaluation of \( \Delta \hat{\phi}_w(t_1, t_2) \) by equation 2-49 is typically very efficient. If we consider the popular four-frame phase shifting algorithm described in section 2.2.1 for example, then computation is simplified by substituting

\[ N(t) = I(3) - I(1), \quad \text{and} \quad D(t) = I(0) - I(2). \]  

(2-52)

Finally, substituting equation 2-49 into equation 2-43 gives:

\[ \Delta \hat{\phi}(t, 0) = \sum_{t'=1}^{t} \tan^{-1}\left\{ \frac{N(t')D(t' - 1) - D(t')N(t' - 1)}{N(t')N(t' - 1) + D(t')D(t' - 1)} \right\} \]  

(2-53)

The main advantage offered by the new algorithm is that spatial smoothing can now be incorporated relatively easily. The two additional computed images \( F(t_1, t_2) \) and \( G(t_1, t_2) \) corresponding to the numerator and denominator in equation 2-53 are proportional to the sine and cosine of the unwrapped phase change respectively, and are thus periodic in nature. Since these images are functions of phase-change values (and not the phase values themselves), random variations in the initial speckle phase from pixel to pixel cancel out. Any remaining discontinuities in these smoothly varying images must therefore be due to physical discontinuities present in the test surface. Simple local averaging operators such as the low pass filter can therefore be applied to these images and provide an effective method of spatially smoothing the phase difference maps (see section 6.2.1 for a discussion of optimal filters). However, one drawback of this approach is that intermediate phase values can no longer be guaranteed to cancel out when spatial smoothing is applied in this way. Measurement errors in these intermediate values can therefore contribute significantly to the final unwrapped phase estimator in the presence of speckle decorrelation as demonstrated experimentally in reference [97].
2.4.3 Modified sum of differences method

The modified sum of difference algorithm proposed by Huntley et al.[97] calculates \( \Delta \hat{\phi}_w(t,0) \) by using equation 2-49 and applying spatial smoothing to the numerator and denominator as discussed in the previous section. A modified form of equation 2-38 is then used to determine the number of \( 2\pi \) phase jumps between two successive measurements of the phase as follows:

\[
d(t) = \text{NINT} \left\{ \Delta \hat{\phi}_w(t,0) - \Delta \hat{\phi}_w(t-1,0) \right\} / 2\pi, \quad t = 2, 3, \ldots, N_i - 1
\]

The total number of phase jumps, \( v(t) \), is now calculated as

\[
v(t) = \sum_{t'=2}^{t} d(t'), \quad t = 2, 3, \ldots, N_i - 1
\]

\[
v(1) = 0
\]

and the unwrapped phase difference obtained as:

\[
\Delta \hat{\phi}(t,0) = \Delta \hat{\phi}_w(t,0) - 2\pi v(t), \quad t = 1, 2, \ldots, N_i - 1
\]

The key difference between this approach and basic algorithm described in section 2.4.1 is that we are now working with wrapped phase-change values rather than with the wrapped phase values themselves. As stated above, this is attractive when applying spatial smoothing to speckle interferometry data where speckle phase can vary randomly from pixel to pixel. The effect of calculating the phase-change values with respect to the phase estimate at time \( t = 0 \) is that intermediate phase values only contribute to the calculation of the integer valued correcting field \( v(t) \).

2.5 Spatial phase unwrapping

Spatial phase unwrapping refers to the process of calculating the integer valued correcting field \( v(m,n,t) \) by comparing adjacent pixels or pixel regions along the spatial axes. In the simplest implementation, the one-dimensional temporal analysis described in the preceding sections can be modified by substituting the time index \( t \) with the spatial index \( m \) and applied by scanning along a single row in the wrapped phase map. The relative phase of each row is then be adjusted by unwrapping along a single column[112]. However, this scheme uses only one of a large number of unwrapping paths available in a two-dimensional image. Unwrapping first the columns then the rows, for example, would be equally acceptable. In fact, all...
unwrapping paths will produce the same result provided that the phase map had been spatially sampled in accordance with the Shannon sampling theorem. In practice, however, this condition is rarely achieved across the whole image and aliasing-induced inconsistencies are generated. Physical discontinuities on the test surface, noise from electronic sources, and steep phase-gradients due to speckle decorrelation can all lead to under-sampling of the phase data. Consequently, significant effort has been applied to develop algorithms that are able to successfully perform spatial unwrapping of a single wrapped phase image containing localised discontinuities. A large number of such algorithms that are now available can be classified as either path-dependent (for example, branch cut methods, tile processing, and cellular automata) or path-independent (for example, minimum $L^p$-norm methods) and are summarised in sections 2.5.1 and 2.5.2 respectively.

In general, spatial unwrapping is applied to $\Delta \hat{\phi}_w (t,0)$, namely the rewrapped difference between the wrapped phase values corresponding to the initial and final states of a test surface. This step is particularly important when applying spatial smoothing to speckle interferometry data where speckle phase can vary randomly from pixel to pixel, and can be implemented using equation 2-49. Alternatively, spatial unwrapping can also be applied directly to subtraction (or addition) correlation fringes but typically produces much poorer results. This ability to generate unwrapped phase maps from only a single wrapped phase (or more usually, phase-change) image has historically made spatial phase unwrapping the technique of choice as it greatly reduces the data sampling rate and data storage requirements of practical systems when compared to those based on temporal phase unwrapping.

The common shortcoming of all spatial unwrapping algorithms is the possibility of propagation of unwrapping errors within the image due to the underlying principal of spatial comparison of phase values. Consequently, erroneous phase values due, for example, to poor modulation, may corrupt valid phase values in other areas of the image. As a result, identifying erroneous pixels and filtering of the wrapped phase map prior to unwrapping becomes an important part of the analysis procedure. An image mask, generated by thresholding a white light image of the test surface over a dark-background, for example, is often used to identify surface boundaries and those
regions with low reflectivity. This masking process often produces isolated regions within the image that can be unwrapped successfully, however, the unconnected zones do not share a common phase origin. Manual intervention is therefore required to produce a consistent solution.

2.5.1 Path-dependent methods

Path-dependent spatial unwrapping algorithms attempt to calculate the integer valued correcting field \( \nu(m,n) \) by identifying an optimum unwrapping path through the image. The presence of pixels with erroneous phase values will cause the calculated integer valued correcting term \( \nu(m,n) \) to become dependent on the unwrapping path chosen. In figure 2-4, for example, a spatial unwrapping algorithm following path 1 from point \( A \) to point \( B \) will detect two phase discontinuities and hence increment \( \nu \) by 2. However, the same algorithm unwrapping along path 2 will only detect a single phase discontinuity and increment \( \nu \) by 1. The resulting unwrapped phase difference between point \( A \) and point \( B \) will therefore differ by \( 2\pi \) depending on the unwrapping path. A common approach used to automatically identify those pixels with an erroneous phase value involves calculating \( \nu \) for a small closed-path around every group of four adjacent pixels. If the phase values along the path are sampled in accordance to the Shannon sampling theorem then \( \nu \) should be zero; a non-zero value indicates that the integration path contains a discontinuity source or residue. Residues produced by speckle decorrelation or electronic noise typically appear in pairs of opposite sign (referred to as dipoles), however, they may also appear individually at surface boundaries (referred to as monopoles). The presence of one or more residues within the wrapped phase map means that the unwrapped phase cannot be determined unambiguously; an infinite number of possible unwrapped phase values can be obtained by integrating along a path that encircles a residue a sufficient number of times. Many methods have been suggested to balance these residues and select the unwrapping path within a wrapped phase map. It should be noted, however, that if residues are present in the wrapped phase map then it is not possible from the phase values alone to determine any definitive path that is guaranteed to avoid unwrapping errors. So called quality maps[125], derived from fringe modulation data for example.
can be used to provide additional information to aid with the selection of an optimal path.

One common method of restoring a single-valued field, and therefore obtaining the unwrapped phase unambiguously, involves placing cut-lines between dipoles (or monopoles and the boundary) as barriers to the unwrapping path. A phase map containing $N_R$ positive and $N_R$ negative residues allows $N_R!$ possible pairings of the residues of opposite signs, and consequently several criteria have been proposed to select the optimum cut distribution. From a statistical viewpoint, pairing residues to minimise the total cut-length (or more precisely, cut-lengths squared) within the image has been shown to be optimal in certain circumstances[126]; a direct implementation that compares all possible pairings, however, is impractical for real data. The “nearest neighbour” algorithm[127] offers a simple approximation and pairs two residues (or one residue and the boundary) that are separated by the shortest distance. However, this simple criterion can fail when the separation between two adjacent dipoles (or a dipole and the boundary) is shorter than the spacing between the component residues.

The “modified nearest neighbour” algorithm[128] is derived from the previous method and can be more reliable when high concentrations of residues are present. The main difference here is that all residues that are likely to constitute a dipole are initially grouped together; simple rules are then applied to split the groups to leave clusters of individual dipoles. However, both of these approaches rely on local search methods that can result in long and physically unlikely branch cuts. Furthermore, although the solution will converge on local minimum cut lengths, the minimum total cut length is not guaranteed.

An alternative algorithm based on graph-theory, the so-called “minimum cost-matching” algorithm[129], is guaranteed to find the global minimum cut-length even for high concentrations of residues. In practice, however, the algorithm requires lengthy processing times and may not converge on a solution. One of the main causes of non-convergence lies in the specification of boundaries to regions of continuous phase. Although thresholding methods can be applied to obtain physical surface boundaries, boundaries to continuous phase can also be present along crack lines and
disbands running through the centre of the surface and are much more difficult to identify.

An alternative unwrapping approach involves segmenting the wrapped phase map into small rectangular regions, or tiles. Once each tile has been unwrapped individually (by using one of the above algorithms, for example), they are mutually phase shifted in order to reduce any phase inconsistencies present at their edges. Tiles with good agreement at their edges are then merged together to form larger tiles in an iterative procedure until the whole image is reconciled. Various criteria have been used to decide which tiles should be merged first in order to maximise confidence in the unwrapped phase. One approach using minimum spanning trees[130] uses a weighting mechanism that is robust both in dealing with spike noise at the pixel level and dealing with phase inconsistencies due to aliasing and surface boundaries at higher levels.

Cellular automata[131] have also been proposed for spatial unwrapping and consist of simple, discrete mathematical systems. The state of each cell (consisting of a 2 × 2 pixel array, for example) evolves in discrete time steps according to simple local neighbourhood rules. Global behaviour is performed by an iterative voting strategy and many iterations a usually required to achieve convergence, however, successful unwrapping is again dependent on a suitable choice of unwrapping boundaries.

2.5.2 Global methods
A second class of spatial unwrapping methods takes a global “fitting” approach to determine the optimum unwrapped phase solution. These algorithms minimise the difference $\varepsilon^p$ between the gradients of the wrapped phase and the solution using the so-called $L^p$-norm minimisation criterion. For the two-dimensional spatial unwrapping problem this can be stated as:

$$
\varepsilon^p = \sum_{m=0}^{N_x-1} \sum_{n=0}^{N_y-1} \left| \Delta \hat{\phi}(m+1,n) - \Delta \hat{\phi}(m,n) - W[\Delta \hat{\phi}_w(m+1,n) - \Delta \hat{\phi}_w(m,n)] \right|^p
$$

$$
+ \sum_{m=0}^{N_x-1} \sum_{n=0}^{N_y-1} \left| \Delta \hat{\phi}(m,n+1) - \Delta \hat{\phi}(m,n) - W[\Delta \hat{\phi}_w(m,n+1) - \Delta \hat{\phi}_w(m,n)] \right|^p
$$

(2-57)
Quantitative interferogram analysis

The global minimisation approach differs from the path-dependent methods in that it does not directly identify residues present in the wrapped phase map. In the particular case where \( p = 2 \) the minimisation problem is a least-squares optimisation and can be solved by application of the discrete cosine transform or fast Fourier transform. However, pixels with erroneous-phase values can corrupt the unwrapped phase map on a global scale unless pixel weighting is applied. Simple binary weightings (0 or 1) may be used, or alternatively, a continuous weighting distribution based on the inverse probability of phase error. Reference [132] provides several examples to demonstrate the performance of the least squares algorithm in the presence of phase errors from various sources.

2.6 Summary

Phase shifting is a popular technique used to obtain the interference phase from intensity distributions as an intermediate step toward obtaining the required physical quantity. This chapter has introduced several phase shifting methods applied both along the time axis and the spatial axes. Practical implementations of these algorithms in the space-time domain are generally more computationally efficient than their frequency domain counterparts, however, this efficiency brings with it some limitations. If we consider the Fourier-transform representation discussed in section 2.2.4, the phase estimate is obtained by applying a narrow bandpass filter in the frequency domain that isolates the frequency-modulated sideband. Ideally, the frequency domain implementation will apply a bandpass filter with a sharp roll-off characteristic to reduce the possibility of spectral leakage. Unfortunately, the corresponding window function in the space-time domain will have a very large spread, with the number of intensity samples \( M \) approaching infinity. Practical space-time phase shifting algorithms must therefore trade-off the filter function roll-off against the number of intensity samples required. This compromise has been the subject of much research over the last ten years and a family of robust phase shifting algorithms are now available as a standard toolset for use in hostile environments.

The process of phase unwrapping is now an essential step in almost all whole-field optical applications. Historically, the combination of temporal phase shifting with spatial phase unwrapping has proved to be popular; however, the increasing
performance to price ratio of more modern computers is making temporal phase
unwrapping more attractive. Forward-looking optical metrology systems are now
requiring that the unwrapping process is both resilient to errors and completed within
a guaranteed time period of a few seconds or less. Consequently, deterministic
temporal phase unwrapping methods with data-dependent run times may be more
suitable for these systems than iterative spatial unwrapping algorithms.

2.7 Tables

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th>Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>3</td>
<td>(\frac{\sqrt{3}(0,-1,1)}{2,-1,-1})</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>4</td>
<td>(0,-1,0,1)</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>5</td>
<td>(0,-2,0,2,0)</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>5</td>
<td>(0.95,-0.59,0.59,0.95)</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>7</td>
<td>(0.95,-0.95,-0.99,0.87,-0.59,-0.21,0.21,...)</td>
</tr>
<tr>
<td>6</td>
<td>15</td>
<td>15</td>
<td>(0.91,0.84,0.31,-0.10,-0.50,-0.81,-0.98,...)</td>
</tr>
</tbody>
</table>

Table 2-1: Examples of sampling coefficients calculated using equations 2-29 and 2-30 for common
phase-shifting algorithms. The diagram indicates the location of phase sampling points on a complex
unit circle. The phase estimate is calculated at the time \(t = 0\). Sampling coefficients are given in the
form \(b(0), b(1),..., b(M-1) / a(0), a(1),..., a(M-1)\).
2.8 Figures

Figure 2-1: (a) The wrapped phase $\Delta \phi_w(t,0)$ obtained using projected fringes to measure the surface profile of two spatially separated turbine blades. (b) The unwrapped phase $\Delta \hat{\phi}(t,0)$ obtained using temporal phase unwrapping. The pseudo-colour map used to display the unwrapped phase represents small values with dark red, ranging though orange, and high values with light yellow. The unwrapped phase values are directly proportional to surface height, with the pseudo-colour map representing a dynamic range of approximately 50 mm. Note that the relative phase origin of each blade can be determined unambiguously even though no noise-free spatial unwrapping path is available. The phase data was obtained using the four-frame phase shifting algorithm and projecting a maximum of 16 fringes across the field of view.
Figure 2-2: Phase unwrapping along the time axis. The unwrapped phase $\hat{\phi}(t,0)$ is obtained from the wrapped phase sequence $\hat{\phi}_w(t)$ by the addition of integer multiples of $2\pi$.

(a)

(b)

Figure 2-3: (a) Fourier transform of continuous time-varying cosine intensity signal with temporal carrier using a rectangular window function. (b) Fourier transform of sampled time-varying cosine intensity signal with temporal carrier using a rectangular window function.
Figure 2-4: Wrapped phase $\Delta\phi_w(t,0)$ obtained using a speckle interferometer to measure out-of-plane displacement of a carbon fibre panel. Erroneous phase values within the image cause the unwrapped phase to become path-dependent. A spatial unwrapping algorithm following path 1 will detect two phase discontinuities, whilst the same algorithm following path 2 will only detect a single phase discontinuity. The resulting unwrapped phase difference between point A and point B will therefore differ by $2\pi$. 
Section II: Real-time temporal phase unwrapping algorithm
3 Real-time digital implementation of the temporal phase unwrapping algorithm

3.1 Introduction

The previous chapter described the temporal phase unwrapping procedure and introduced three methods that may be used to compute the unwrapped phase. We now concentrate on the modified sum of differences method and investigate the various implementation considerations that arise when realising a practical system. In particular, we are specifically interested in selecting a suitable development platform that can be used to develop digital temporal phase unwrapping systems tailored for real-time high-speed profilometry and digital speckle pattern interferometry (DSPI) applications.

The set of algorithms that make up an application is typically developed and proven on computers with no constraints on execution time, size, or cost, since the purpose of the research environment is to provide maximum flexibility and observability to the signal processing researcher. Conversely, a practical deployment system must produce its results in real-time, economically, and within the limited volume and power that allow the system to be moved to the signal source (rather than delivering the signal to a massive immobile processor).

Preparing the algorithm for deployment requires tradeoffs and integration across the multiple domains of algorithm, hardware, and software. Changes to both the structure of the algorithm and the structure of the hardware and software that execute the algorithm are usually needed. Such changes can include introducing more computation steps to achieve regularity of computation, decomposing an algorithm onto a particular architecture, or devising a custom architecture suited for the algorithm.

It is a common misconception that image processing in real-time is no more than a minor variation on image processing without regard to time, i.e. that faster machines will make any speed problems eventually go away. But real-time imaging is not just...
Real-time digital implementation of the temporal phase unwrapping algorithm

about speedy hardware. Real-time image processing involves at least three fundamental trade-offs: (a) performance versus image resolution; (b) performance versus storage and input/output bandwidth; and (c) the number of tasks versus synchronisation. Of these problems, faster machines could solve only the first and possibly the second.

Real-time operation requires that the time between the presentation of a set of inputs and the appearance of all the associated outputs (the so-called response time) is explicitly bounded. Equivalently, a real-time system is one whose logical correctness is based both on the correctness of the outputs and their timeliness. The timeliness constraints or deadlines are a reflection of the underlying physical process and method of acquisition. The deadlines in the proposed digital temporal phase unwrapping system are determined by the need to process phase-shifted interferograms received from the camera and generate continuous unwrapped phase display updates at a constant rate.

The complexity of a real-time implementation increases with data throughput and the number of processing operations that must be performed on this data. It is useful therefore to estimate the data throughput required for a practical real-time temporal phase unwrapping system. The data throughput can be calculated as the product of the spatial resolution and temporal resolution; if we assume that each interferogram consists of at least $512 \times 512$ pixels recorded by the photodetector array, and that 60 such interferograms are recorded per second, then the minimum input data rate is $512 \times 512 \times 60 = 16 \times 10^6$ intensity samples per second. Assuming that the system implements a four-frame phase stepping algorithm to estimate the wrapped phase values, $4 \times 10^6$ phase values must be unwrapped per second to achieve continuous real-time operation.

3.2 Temporal phase unwrapping algorithm

We now look at the specific considerations required to implement a real-time digital system designed to both: (a) maintain a real-time up-to-date estimate of the wrapped phase calculated from a time series of phase shifted intensity samples; and (b) maintain an up-to-date estimate of the unwrapped phase using the temporal phase
Real-time digital implementation of the temporal phase unwrapping algorithm
unwrapping algorithm. The modified sum of differences method of temporal phase
unwrapping (described in section 2.4.3) has been chosen for this study since it is
easily adapted to spatial speckle averaging, and is insensitive to small errors in the
intermediate phase values that only contribute to the calculation of the integer valued
correcting field. The phase values computed by the real-time system will typically
correspond to the change in some physical parameter between the initial condition
when the system is started \((t = 0)\) and the current time \((t)\). Practical applications of
the system will use a CCD camera to acquire phase shifted two-dimensional intensity
images. The system will therefore be required to maintain this data format throughout,
taking two-dimensional intensity data at the input and presenting two-dimensional
wrapped and unwrapped phase estimate data sets at the output.

The number of samples required to begin processing will depend directly on the
choice of phase shifting scheme; the processing frame is therefore equal to the
number of intensity samples, \(M\), required by the phase shifting algorithm. We
therefore define the system input as a vector of phase shifted intensity samples:
\[
x(t) = [I(0), I(1), I(2), \ldots, I(M-1)]
\]
(3-1)
such that a new phase estimate can be computed for each \(x(t)\) input received.

A simplified high-level schematic representation of the digital system in which the
spatial dependence of the data sets has been dropped for clarity is shown in figure 3-1.
The data processing has been partitioned to enable the performance at each stage to be
studied independently for various implementations: stage one takes the input vector
\(x(t)\) and computes the \(F(t, 0)\) and \(G(t, 0)\) images that are proportional to the sine and
cosine of the phase change respectively; stage two performs spatial smoothing of
these images (as usually required in speckle applications); stage three implements the
arctangent function and outputs the wrapped phase change estimate \(\Delta \hat{\phi}_w(t, 0)\); stage
four implements temporal phase unwrapping and presents the unwrapped phase
change estimate \(\Delta \hat{\phi}(t, 0)\) at the output. The following paragraphs present a brief
discussion of the required processing steps and explain the partitioning scheme.

The first stage involves demodulating the phase stepped intensity samples to compute
the \(N(t)\) and \(D(t)\) signals according to equations 2-27 and 2-28 respectively. Since
the sampling coefficients $a(t)$ and $b(t)$ are specific to the phase shifting algorithm
the implementation of the decoding block is dependent on the phase shifting scheme.
Subsequent processing of the $N(t)$ and $D(t)$ signals, however, is the same for all
phase shifting algorithms and one could consider treating the decoding block as a
discrete processing stage. Instead, the approach taken in this study is to include the
complex division operation described by equations 2-50 and 2-51 within the first stage.
The main consideration here is that there is no spatial dependence in the data
processing up to this point and thus $F(t,0)$ and $G(t,0)$ can be computed without
regard to the spatial position of the current sample within the two-dimensional data
set. In contrast the following spatial smoothing operation is strongly spatially
dependent and provides a natural break in data flow.

The second stage performs spatial smoothing of the $F(t,0)$ and $G(t,0)$ images to
generate $F'(t,0)$ and $G'(t,0)$ respectively. Convolution of the two-dimensional data
sets with a small two-dimensional kernel (typically $3 \times 3$ pixels) involves
neighbourhood multiply and accumulate (NMAC) operations and is the only part of
the system that utilises the spatial position of data samples. Since the smoothing
operation is only required when working on speckle interferograms it is useful to
isolate this process so that performance figures relating to the other processing steps
remain valid when smoothing is omitted.

The third processing stage computes the wrapped phase change estimate $\Delta \phi_w(t,0)$
using the four-quadrant arctangent function, denoted by $\text{atan2}(y/x)$. Implementation of
trigonometric functions in digital systems is always problematic due to their non-
linear nature and typically involves the use of a lookup table (LUT) or polynomial
series expansion. Lookup tables minimise run-time computation[133] by storing pre-
calculated results for a finite set of operands; the memory required to store these
results depends both on the operand resolution and output resolution. On the other
hand, polynomial expansion offers the advantage of increasing precision simply by
including higher order terms in the expansion resulting in a corresponding increase in
run-time computation. The precision of the power series approximation typically
varies over the input domain and often requires the input range to be controlled to
avoid unstable regions. The four-quadrant arctangent function $\text{atan2}(y/x)$ can be
Real-time digital implementation of the temporal phase unwrapping algorithm implemented as a single-quadrant arctangent function \( \text{atan}(z) \) (where \( z = y/x \)) followed by rotation to the correct quadrant. Abramowitz et al. \cite{134} provide several polynomial approximations to the arctangent function enabling the computational complexity to be balanced against the required precision. The Chebyshev (also known as “min-max”) condition \cite{135} can be used to obtain the power series 
\[ p(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n \]
by minimising the peak absolute error 
\[ e(z) = \text{atan}(z) - p(z) \] for a given value of \( n \), for example. For \( n = 3 \), the coefficients 
\[ a_1 = 0.97239199564847 \quad \text{and} \quad a_3 = -0.19194538972643 \]
result in \( e(z) < 0.004952 \) radians for \( |z| \leq 1 \). The restricted input range \( |z| \leq 1 \) can be overcome by using the trigonometric identity 
\[ \text{atan}(y/x) = (\pi/2) - \text{atan}(x/y) \] Alternatively, a less computationally intensive heuristic approximation can be defined for all four quadrants of the input domain:

\[
\begin{align*}
\text{atan} 2\left(\frac{y}{x}\right) &= r, \quad \text{if} \quad x \geq 0 \\
\text{atan} 2\left(\frac{y}{x}\right) &= -r, \quad \text{otherwise}
\end{align*}
\]
where

\[
\begin{align*}
r &= \frac{\pi}{4} \left(1 - \frac{|x| + y}{|x| + y}\right), \quad \text{if} \quad y \geq 0 \\
r &= \frac{\pi}{4} \left(3 - \frac{|x| + y}{|x| - y}\right), \quad \text{otherwise}
\end{align*}
\]
However, analysis of this approach reveals that the precision has been compromised for efficiency, with \( |e(z)| < 0.07 \) radians. Evaluation of the arctangent function is therefore investigated separately in stage three to allow the performance of various implementations to be compared.

Finally, stage four implements the phase unwrapping procedure following equations 2-54 through 2-56. The \( Z^{-1} \) operation denotes a delay of one sample period, implementation of which requires the storage and subsequent readout of one complete two-dimensional data set. The function \( \text{NINT}(x) \) rounds to nearest integer; halfway cases are rounded to the integer value larger in magnitude (this corresponds to the
Real-time digital implementation of the temporal phase unwrapping algorithm

Fortran intrinsic function NINT(x) and is therefore often referred to as Fortran rounding). Itoh\[123\] noted as far back as 1982 that efficient nearest integer rounding could be achieved using a 2s-complement fixed-point number representation and simply truncating the operand.

### 3.3 Implementation considerations

#### 3.3.1 Algorithm design

In selecting an algorithm and implementation some consideration needs to be given to the problem requirements. For example, how often is the application to be used? If an algorithm is to be run only a few times on cases that are not too large, then it is certainly preferable to have the computer take a little extra time running a slightly less efficient algorithm than to have the designer take a significant amount of extra time developing a sophisticated implementation. In the case of the proposed real-time temporal phase unwrapping application, however, two requirements identify the need for an optimised implementation. First and foremost, the system is to operate in real-time, therefore necessitating that the computations on the current processing frame are completed before the arrival of the next. Secondly, timely completion of the required processing steps on the very large data sets that result from the proposed spatial and temporal resolutions is only likely to be achieved through algorithm optimisation; a brute-force approach will not achieve the desired throughput even with today’s high-performance workstation-class processors.

It is often the case that mathematical analysis can shed very little light on how well a given algorithm can be expected to perform in a given situation. In such cases, we need to rely on empirical analysis, where the algorithm is carefully implemented and its performance monitored on “typical” input. A comparison of run-times of the available algorithms can help in understanding their properties. An excellent way to check the efficacy of a particular modification or implementation idea is to run alternative versions on the same input data, then pay more attention to the faster one.

All algorithms must be reduced to data and instruction flows through logic circuits, and this requirement restricts the mathematical structures that are directly appropriate to digital processing. Linear filters involve addition and multiplication and can often
Real-time digital implementation of the temporal phase unwrapping algorithm involve real numbers as well as integers. Both addition and multiplication of floating-point numbers (see appendix A4) take significantly longer than their integer counterparts because these operations must be implemented via a firmware subroutine involving multiple integer shifts, adds, and multiplies. Although modern arithmetic logic units (ALUs) increasingly support floating-point operation directly in microcode these operations are usually still slower than the corresponding integer operations. Conversely, non-linear filters only require integer (or binary) operations, typically involving subtraction and logical operators (e.g. AND, OR, MAX, MIN), but not multiplication. These operations are generally among the fastest macroinstructions provided in the ALU and therefore ideal for many real-time applications.

Clearly, the issues under discussion here are extremely system- and machine-dependent. A rather detailed knowledge of the hardware architecture issues, operating system and programming environment are therefore required when embarking on a serious attempt to optimise the design of a digital system.

3.3.2 Hardware architectures

Ultimately the sophistication of any data processing system will depend to a large extent on the electronic hardware on which it is implemented. In turn the choice of hardware is dependent both on the time and money available to design, prototype, manufacture, and integrate custom devices. In most instances the integration level at which the hardware design is to be specified must be made early in the project. At the lowest level, very high speed processing systems can be constructed from discrete digital and analogue electronic devices. This approach has the advantage of absolute control over the choice of devices available to the system and the routing of signals between those devices. In 1983, for example, Mertz[136] described a real-time fringe analysis system able to process an analogue video signal using a discrete analogue-to-digital arctangent converter at a rate of 5x10^6 samples s^{-1}. However, implementing more than the simplest of algorithms in this way soon becomes very cumbersome.

A more flexible approach can be afforded using custom designed application specific integrated circuit (ASIC) or field programmable gate array (FPGA) devices which offer the equivalent of many thousands of discrete devices (or processing elements)
Real-time digital implementation of the temporal phase unwrapping algorithm within a single silicon die. Typically, the on-chip interconnection between these devices can be simulated in software before the final design is implemented in silicon. However, whilst custom ASICs enable the end-user to specify which processing elements should be included on the die, prices are prohibitive unless ordering in large quantities. Low cost FPGAs on the other hand are available off the self and provide a fixed collection of processing elements on a single die; the algorithm must therefore be mapped to those processing elements that are available. Interfacing to real-world peripherals (e.g. sensors, cameras, etc.) can be particularly difficult when working with ASIC and FPGA implementations. Typically, the low-level interfacing logic must be “hard-coded” into the design, either by including propriety intellectual property in the chip design or implementing custom interfaces; the result usually offers limited compatibility with alternative sensors. In particular, interfacing to modern digital cameras is often not straightforward with many manufacturers implementing non-standard protocols.

General-purpose microcontrollers and microprocessors have proved a popular choice for many custom system designs, perhaps due to their low-cost, availability, and well-documented industrial standard architectures. Many manufacturers now supply products ranging from highly-integrated low-performance controllers through to high-performance workstation-class processors enabling system designers to select devices which closely meet their needs. However, the data throughput requirements for the proposed system reduce the choice of suitable devices to the high-performance central processing units (CPUs) typically found in personal computers and workstations.

The CPU functionality can typically be divided into a control unit and a processing unit, although the boundary is not always clear. The control unit is responsible for fetching instructions and operands from memory, storing results to memory, and computing the address of the next instruction. The processing unit is able to perform logical operations on individual bits and arithmetic on word-based data values, and therefore is often referred to as the arithmetic logic unit, or ALU. Arithmetic instructions may require a varying number of clock cycles to complete. Simple integer addition, subtraction and compare operations are among the fastest to execute, however, multiplication, division, and floating point arithmetic can require multiple instruction cycles to complete.
Real-time digital implementation of the temporal phase unwrapping algorithm

The von Neumann architecture, typical of general-purpose microprocessors, uses the same memory for both program and data. Program and data information are usually segregated into different address regions within the memory. Processor memory is arranged in a hierarchical manner: a small number of memory registers (typically in the region of 32) are available directly to the ALU; a local high-speed memory cache contains the most recently accessed instructions and data from main memory; a small amount of general purpose memory is often available on-chip; memory available off-chip is not constrained by chip capacity, but typically has longer access times than memory located on-chip.

Performance gains have been achieved to some extent by increasing the CPU clock rate made possible by advances in integrated circuit technology allowing more dense transistor placement, shorter propagation times, and the ability to move more logic on-chip. Further performance gain has arisen from architectural enhancements that allow the simultaneous processing of multiple instructions and cascading of the multiple steps (pre-fetch, instruction decode, operand fetch, execute, write-back) required to execute an instruction in a manufacturing assembly line fashion (so-called pipelining). Increased cache memory sizes are often used to move relevant data and instructions closer to the processor where they may be accessed more quickly. Increased data path widths now enable more bits to be transferred in parallel, yet the data itself need only be sampled at a fixed data width. Features have now been added to modern CPUs to allow the wide data path to be divided into parallel fields to handle multiple samples of lower-bit widths. The architecture is enhanced with new single instruction multiple data (SIMD) instructions that allow the same instruction to operate on multiple sub-word data values. The increased complexity associated with improved performance is illustrated by the 700-fold increase in the number of transistors used in the Intel Pentium II compared to the original Intel 8086. Fleischer et al.[137] have taken advantage of SIMD instructions on a standard 266 MHz Intel Pentium II microprocessor to implement a three-dimensional confocal microscopy profilometer that is able to process $768 \times 576 \times 25 = 11.1 \times 10^6$ intensity samples s$^{-1}$.

The general-purpose microprocessor is designed to handle wide-ranging mixtures of operations and control flow that can be data-dependent, making large jumps from one area of program memory to another. Conversely, traditional digital signal processing
Real-time digital implementation of the temporal phase unwrapping algorithm typically emphasises repetitive multiply-and-add (or multiply-accumulate) operations that step sequentially through data values stored in consecutive memory locations. Dedicated digital signal processors (DSPs) that exploit these properties can therefore be used to increase data throughput in many digital systems.

DSP architectures are optimised to improve the performance of the more specialised signal processing operations. The ALU elements typically feed directly into fully parallel hardwired multiplier elements and adders to facilitate multiply-accumulate operations. The DSP instruction set is designed so that all instructions take an equal (and minimal) number of clock cycles; equal-period instructions then allow operations on separate processing units to be cascaded in a pipeline fashion for improved throughput.

In contrast to the single-memory von Neumann architecture associated with general-purpose microprocessors, DSP systems often use the Harvard architecture in which data and instructions are stored in separate memories that can be accessed in parallel using dedicated buses. Further speed is introduced into the multiply-accumulate pipeline by providing separate memories and associated buses for each of the multiplier inputs, allowing both operands to be fetched from memory in a single instruction cycle.

The Analog Devices ADSP-2106x, also known as the Super Harvard Architecture Computer (SHARC), is an example of one such high-performance floating-point DSP. The core contains three independent parallel floating-point units (containing multiplier, barrel shifter, and ALU) and allows three instructions to be executed together in each cycle. Two large dual-port memories and associated buses are provided on-chip to supply data to each of the multiplier inputs. In addition, the SHARC DSP supports multiprocessing, allowing as many as six processors to be combined without additional interface circuitry, and enables data to be transferred between devices. Gdeisat et al. [138] describe a real-time spatial fringe analysis system based on linear digital phase locked loop implemented using a pipeline of six 40 MHz SHARC processors that enables continuous phase maps to unwrapped at $256 \times 256 \times 25 = 1.6 \times 10^6$ samples s$^{-1}$. 

83
The conventional dividing lines between these hardware technologies are increasingly becoming blurred with manufacturers offering off-the-shelf modular integrated products that incorporate both sophisticated high-level processors and low-level programmable FPGAs. Many such products are available in industry standard PCI and VME formats enabling custom architectures to interface directly with low cost personal computers and workstations. Such products bridge the gap between dedicated custom hardware solutions and computers based on general-purpose processors. In addition, such products often provide support for interfacing to a wide range of sensors and cameras. Many such examples can be found; the following are chosen only to demonstrate the range of choice available. Kane Computing’s Multi PCI card product range integrates multiple TMS320 digital signal processors, pixel processing FPGA elements (LUTs, non-linear ALUs, etc) and camera interfacing logic. Coreco’s Viper PCI card product range integrates end-user programmable FPGAs to perform pixel processing and camera interfacing logic. Coreco’s alternative Mamba product range incorporates a complete Intel Pentium II system (together with its own embedded operating system stored on a disk-on-chip device) and camera interfacing logic on a single PCI card. Datacube’s MaxVision and Image Technology’s MVC150 PCI and VME pipeline processing product ranges integrate proprietary ASICs, FPGAs and camera interfacing logic onto a single board; the end-user then specifies the interconnections between the image processing elements (LUTs, ALUs, NMACs, etc.) to form sequences of processing operations (rather like a manufacturing assembly line) that can be reorganised dynamically between data transfers.

3.3.3 Software issues
Software controlling the hardware processing elements discussed in the previous section can directly affect system performance. Efficient use of the available hardware resources is essential if computationally intensive digital systems are to be implemented in a cost effective manner. Where the algorithm to hardware mapping is not directly 1:1, several computation steps must be performed on a single hardware device in a time-sharing fashion. Allocation and scheduling of hardware resources on complex microprocessor architectures is usually performed by a software operating system. General-purpose operating systems (e.g. Microsoft Windows and Sun Solaris)
Real-time digital implementation of the temporal phase unwrapping algorithm are designed to allocate the hardware on an equal time-sharing basis between several user and system tasks (e.g. word-processing, file caching, etc), often denying the software programmer direct control of the hardware. However, the operating system requirements of real-time digital systems are somewhat different; computation tasks need to be prioritised so that important time critical events (e.g. actuator control) can be serviced immediately without waiting for less important tasks (e.g. keyboard responses) to complete. Consequently, several industrial strength real-time operating systems are now available that enable the response time of specific tasks to be both deterministic and constrained.

The required precision and range of numeric data manipulated by the digital system can impact the complexity of both software and hardware. Allocating more bits to represent numeric values increases the number of quantisation levels available to sample the continuous numeric space but typically also increases the implementation complexity. As the number of arithmetic operations increase, the need for bits increases. For example, adding two $N$-bit numbers results in a sum of $N+1$ bits, and multiplying two $N$-bit numbers yields a product of $2N$ bits. Hardware processors (with the exception of bit serial architectures) have a fixed number of bits in the arithmetic path and consequently results that exceed the allowed number of bits must be reduced in size, either by scaling, truncation, or rounding. Numerical values can be represented using either a fixed-point or floating-point scheme (see appendix A4). The floating-point representation, which reserves a sub-field to serve as a scaling component, increases dynamic range at the expense of added complexity of floating-point arithmetic. The finite word length imposed on both floating-point and fixed-point representations inevitably introduces errors due to truncation and/or rounding of the result. There is therefore a compromise between numerical precision and complexity that must be balanced in the design phase.

Data intensive algorithms (including most image processing applications) implemented using general-purpose microprocessors can benefit significantly from optimising the sequence of memory accesses. In particular, the performance of processing architectures that provide a memory cache is affected by the alignment of data stored in external memory and on the processor stack. When a memory read operation from a cacheable address misses the data cache, the entire line is brought
into the cache from external memory (a line fill). On Intel Pentium (P5) and dynamic execution processors (Pentium II onwards), for example, data arrives in a burst composed of four 8-byte sections to match the cache line size of 32 bytes. With the exception of the Pentium, these processors also fetch the entire 32-byte cache line when a write operation results in a cache miss. Since the cache lines are always aligned to a physical address divisible by 32, aligning of data on 32-byte boundaries in memory therefore enables one to take advantage of matching the cache line size and avoiding multiple memory-cache line transfers (requiring up to 10 clock cycles). Whilst some advanced compilers allow the user to specify the alignment of data stored in main memory, data padding may also be required to optimise the data stack in tight inner loops. This is particularly important as compilers allocate all data that is not static on the stack as a matter of convention.

In digital image processing systems the so-called double buffering (or ping-pong buffering) technique is particularly important to achieve efficient use of the hardware, allowing one half of a memory array to be filled with data associated with one analysis block while the other half, containing data for the previous analysis block, is emptied. This multiplexing is useful for two processors connected in series as it helps to avoid stalling the computation by decoupling the data flow. While the upstream processor produces results \( x_n(m), \ m = 0, 1, \ldots, L - 1 \) for analysis frame \( n \), the downstream processor begins the next stage of processing by ingesting \( x_{n-1}(m) \), which are the results that were produced by the upstream processor during the previous analysis frame. The requirement to continually receive consecutive images from a camera whilst concurrently processing the acquired images provides one example where double buffering is beneficial since the upstream process cannot afford to wait for the downstream process to complete.

3.3.4 Summary

Whilst the mathematical equations presented in chapter 2 precisely describe the computation steps required to perform temporal phase unwrapping, the preceding sections highlight the many choices that must be made not only in algorithm design, but also the hardware and software domains. The proposed real-time system operation, concurrent image acquisition and processing, large data sets, and high data
throughput constraints indicate that the system design will need to be optimised in each of these domains to achieve the required specification. From the array of hardware platforms presented, two architectures have been chosen for further consideration. The general-purpose von Neumann architecture together with the Intel Pentium II processor (state of the art in Q1 1996 when choosing the development platform) with industry standard PCI interface bus provides a low-cost and flexible solution with standard development tools. Implementing the system using this platform would allow off the shelf PCI sensor and camera interfaces to be used and thus reduce the overall development time. Section 3.4 profiles various implementations of the temporal phase unwrapping system on this platform to assess its viability. Section 3.5 investigates the viability of commercial pipeline image processing architectures, optimised for concurrent image processing tasks, as the development platform. These systems offer modular architectures, deterministic run-times for standard image processing operations, and flexible camera interfacing options.

3.4 Viability of von Neumann architecture implementation

The von Neumann architecture is used almost exclusively in all commercially available personal computers and workstations. Competition between manufacturers enables high-performance hardware and development tools based on this architecture to be obtained at lower prices than corresponding systems based on other architectures.

Figure 3-2 shows a schematic representation of a development platform based on the von Neumann architecture. Mapping the temporal phase unwrapping algorithm to this architecture is relatively straightforward since all computation required for a given processing frame is performed by the general-purpose microprocessor (CPU) and is simply queued until processor time is available. Real-time performance can only be achieved, however, if this queue can be cleared before the arrival of the following frame. It should be noted, that all image data transfers between the CPU, camera, video cards, and system memory must pass through the bridge chipset. Furthermore, there can only be one device writing to, and one device reading from, the PCI bus at any one time. Together these issues create a bottleneck in the flow of data between the
Real-time digital implementation of the temporal phase unwrapping algorithm

various system components with the result that concurrent image acquisition, image
processing, and updates of the display are difficult to achieve on this architecture. The
following paragraphs study only the performance of the image processing problem,
adding various modifications to the basic algorithm to improve data throughput.
Image acquisition and display updates will clearly add further loading to the
microprocessor and bridge chipset with a corresponding reduction in data throughput.

A Dell 333 MHz Pentium II personal computer running the Linux operating system
was used to study nine implementations (denoted by TPUA1 through TPUA9) of the
digital temporal phase unwrapping system described in section 3.2. The standard four-
frame phase stepping algorithm ($N = M = 4$) has been used throughout. A summary
of the main characteristics of each implementation is included in table 3-1. Each
implementation was profiled using the same data set of 400 greyscale images (100
sets of 4 phase shifted interferograms) with a spatial resolution of 239 x 192. The
interferograms were recorded using the high-speed digital speckle pattern
interferometer described by Huntley et al.[97] and describe the deformation of a thin
metal plate subjected to an out-of-plane displacement on the back surface towards the
camera. The time taken to complete each of the four stages of the algorithm was
recorded independently for each processing frame using the Time Stamp Counter
CPU register that increments once every clock cycle.

Each implementation was assigned to the real-time task scheduler at a higher priority
than other applications running in the multi-tasking operating system to ensure that
processor time was not diverted away from the profiled task. The C programming
language was used for the majority of the coding as it enables arithmetic operations to
be expressed at the bit level. Assembly language was used to control stack alignment,
read the Time Stamp Counter register, and implement the SIMD instructions (referred
to as MMX instructions on Intel CPUs). All memory required by the application was
allocated during initialisation; paged memory was then locked into physical RAM to
prevent virtual memory swaps occurring whilst timing. Image data was loaded into
the allocated RAM before starting the profiling operation to avoid filesystem accesses
influencing the results. Image data and stack were aligned on 32-byte address
boundaries to optimise cached memory accesses and inline coding was used to avoid
function calls within the inner loops. The timing results for each implementation are
Real-time digital implementation of the temporal phase unwrapping algorithm included in figure 3-3 and the corresponding data throughput values shown in figure 3-4. Detailed description of the results for each implementation is inappropriate for this thesis and therefore only the main results are presented below.

Implementation TPUA1 is a direct mapping of the algorithm using double (64-bit) precision floating-point number representation and serves as a reference point for the profiling study. High precision is maintained by using floating-point representation at the expense of data throughput. Neighbourhood multiply and accumulate (NMAC) operations are implemented by convolving the operands with a uniform 3×3 pixel kernel. The intrinsic arctangent function atan2() included in the C language math libraries has been used for maximum precision. The final wrapped phase difference map \( \Delta \phi_w(t,0) \) (output from stage 3) is shown in figure 3-5 and a cross-section along row 112 (i.e. 80 rows from the bottom of the image) of the final unwrapped phase difference map \( \Delta \phi(t,0) \) (output from stage 4) is included in figure 3-6. The mean recorded data throughput was approximately \( 0.5 \times 10^6 \) samples s\(^{-1}\) and is around 8 times below that required for the proposed system.

Implementation TPUA2 uses single precision (32-bit) floating point throughput and results in a small improvement in throughput. Figure 3-7 shows the computational error introduced into the unwrapped phase values with respect to TPUA1 after 100 processing frames; numerical errors introduced by the reduced numerical resolution are around \(-150\) dB and can safely be ignored for most applications. In order to achieve further throughput improvements the remaining implementations (TPUA3 through TPUA9) make use of the fixed-point number representation. However, to achieve maximum throughput gains they all restrict the data path width L2 to 8 bits (see figure 3-1). TPUA2 was modified (denoted TPUA2*) to simulate the 8-bit fixed-point path at L2 so that the resulting additional error introduced could be quantified. TPUA2* results are included in figure 3-7 and indicate root mean square (RMS) errors at approximately \(-70\) dB after 100 processing frames. The compromise between increase error load and potential improvement in throughput by the use of 8-bit word-lengths at L2 (through use of SIMD instructions, for example) was seen as justified for the proposed system. The error analysis for implementations TPUA3 through
Real-time digital implementation of the temporal phase unwrapping algorithm

TPUA9 are calculated with respect to TPUA2* in order to assess further additional error introduced by the various fixed-point modifications.

Implementations TPUA3 and TPUA4 use fixed-point computation for stages 1, 2, and 4, but continue to rely on the floating-point intrinsic atan2() function. TPUA3 restricts the data path width L7 to 8 bits in order to reduce the memory required to implement the $Z^{-1}$ operator; TPUA4 increases L7 to 16 bits. Figure 3-8 shows the corresponding errors with respect to TPUA2*; L7 implemented as a 16-bit data path in preference to an 8-bit path significantly reduces additional errors without incurring any noticeable reduction in throughput.

TPUA5 and TPUA6 seek to reduce the time spent computing the arctangent function in stage 3 by using floating-point and fixed-point implementations, respectively, of the heuristic approximation described by equations 3-2 and 3-3. Error analysis results with respect to TPUA2* are shown in figure 3-9 and indicate that both implementations have similar error properties (as one might expect). The fixed-point implementation is somewhat faster, however, and achieves an overall throughput approximately double that achieved using TPUA1.

TPUA7 and TPUA8 seek to improve the throughput further by using a lookup table of pre-calculated results to implement the arctangent function. The 64 kb lookup table takes two 8-bit fixed-point operands and generates a 16-bit fixed-point result. Consequently the results from stage 2 (up to 16 bits) must be truncated to ensure that the LUT operands are within range. TPUA7 applies a fixed truncation scheme for optimal throughput; however, this introduces significant error for poorly modulated pixels. TPUA8 introduces an additional step to normalise the results from stage 2 before truncating in order to retain as many significant bits as possible on a pixel-by-pixel basis. The error analysis results in figure 3-10 clearly show that errors introduced by truncation of the LUT operands have been reduced by approximately 5 times for the speckle interferograms used for this trial simply by introducing the normalisation step; there is, however, a corresponding drop in throughput to be incurred.

TPUA9 modifies TPUA8 using MMX (single instruction multiple data) instructions to implement stages 1 and 2. The decision to restrict the L2 data paths to 8-bit enables
eight operands to be processed in a single instruction cycle resulting in significant throughput gains in stage 1; efficient implementation of the convolution operation using MMX instructions enables similar gains in stage 2. As expected the error analysis for TPUA9 shown in figure 3-10 is almost identical to TPUA8, however, the overall throughput has nearly doubled due to the time-savings made in stages 1 and 2.

In spite of our best efforts, however, the maximum data throughput achieved using the von Neumann architecture described above would still need to be increased by a further 2.5 times to achieve the requirements of the proposed system. Furthermore, the added load of concurrent image acquisition, display updates, and control of external devices (e.g. PZT actuators) will only worsen these figures. Therefore, despite the many advantages and familiarity offered by the von Neumann architecture it was decided that an alternative platform was required.

3.5 Viability of pipeline imaging processing architecture implementation

Commercial pipeline image processors offer flexible hardware architectures and are well suited for implementation for many fringe-processing operations where the same arithmetic operation is required for each pixel in a sequence of images. Flexible architectures enable sequential and parallel data paths to be configured in hardware, with multiple dedicated arithmetic processors placed along each data path as required by the algorithm. Higher data throughput rates than obtained using von Neumann architectures can be achieved by performing concurrent arithmetic operations, however, this must be balanced against the significant development effort required to map a given algorithm to finite hardware resources.

Several research groups have developed fringe-processing systems based on commercial pipeline image processing architectures. In 1989 Stetson et al.[139] and Vrooman et al.[140] independently described speckle fringe analysis systems that use four phase-shifted interferograms to estimate the wrapped phase maps. The Vrooman system uses a pipeline image processing platform to estimate the wrapped phase maps and achieves a throughput of \(1.1 \times 10^6\) wrapped phase samples s\(^{-1}\). A spatial phase unwrapping algorithm optimised for speed and implemented on a general-purpose microprocessor then transforms the wrapped phase maps, requiring approximately
Real-time digital implementation of the temporal phase unwrapping algorithm 2 minutes to unwrap each map. In 1994 van Haasteren et al. [141] described a real-time speckle interferometry fringe analysis system using three phase shifted interferograms acquired simultaneously by three independent cameras at 25 frames s⁻¹. Wrapped phase values are calculated at 10.6×10⁶ samples s⁻¹ and transmitted to a live display; throughput drops by 50% if spatial smoothing is included in the processing pipeline.

Two pipeline processing systems (commercially available in Q1 1996) were considered for the proposed real-time digital system: the Datacube MV250 based MaxTD system and the Imaging Technology MV150/40. Table 3-2 provides a comparison of some of the key features of each system. Both systems interface to a host computer that is responsible for controlling interconnects between processing elements and scheduling data transfers. A schematic representation of the image processing architecture is shown in figure 3-11. In contrast to the bridge chipset used in the von Neumann architecture, the central crosspoint switch enables many inter-device data transfers to proceed concurrently helping to avoid bottlenecks in the data flow. However, one of the main concerns with the MV150/40 system was the reduced number of paths available through the main crosspoint switch compared to the MV250. Both systems support industrial strength real-time operating systems running on the host computer enabling tight control of the task scheduling, however, the cost associated with the MV150/40 VxWorks operating system cross platform development software was approximately one order of magnitude more expensive (at approximately £5k) than the Datacube LynxOs offering. In addition, the self-hosting Unix-like LynxOs operating system significantly simplifies application development by providing industry standard POSIX.1b compliant function libraries and protected virtual memory addressing. Conversely, the relatively simple VxWorks task model has enabled many third party vendors to provide device drivers for peripheral systems that are not available for the LynxOs operating system.

These considerations eventually led to a more detailed analysis of a Datacube MaxTD system (shown in figure 3-12) consisting of a Motorola MVME167 (based on a 33 MHz 68040 CPU) single board computer and two MaxVideo250 (MV250) pipeline processor boards connected across a VME backplane (see appendix A5 for manufacturers’ datasheets). The MV250 cards are additionally connected across a
Real-time digital implementation of the temporal phase unwrapping algorithm dedicated proprietary image data bus (MAXbus) that enables interconnects between processing elements physically located on different cards without loading the general-purpose VME backplane. The system operates at a fixed pixel clock rate of 20 MHz; all processing elements are synchronised to this pixel clock and complete the current pixel processing operation at the same time thus facilitating parallel processing. Numerical processing functions are supported by two ALUs each comprising of a linear and non-linear section (denoted AU_L and AU_N respectively) as shown in figure 3-13. Fixed-point addition, subtraction, and multiplication arithmetic operations are supported, with the notable exception of division. One 512 kb hardware LUT (16-bit operand and 16-bit result) is available that supports four banks of pre-calculated results that can be selected by the CPU. Two other 8 kb hardware LUTs (12-bit operand and 16-bit result) are also available. Hardwired neighbourhood multiply and accumulate operations can be performed on 12-bit operands convolved with a 12-bit kernel (up to 8 × 8 pixels). Twelve independent 4 Mb dual-ported memories (so-called virtual surface image memories or VSIMs) enable double-buffering of data between processing elements. Gateway elements provide flexible control of memory addressing, enabling hardware support for region of interest (ROI) operations, digital image expansion along both dimensions (zooming), and masking. Simultaneous interfacing for two digital cameras and two analogue displays is also included.

In spite of the array of image processing operations supported in hardware, there are several important compromises that have to be made. Mapping of the proposed digital system to the finite set of processing elements described above requires careful design. Fixed word-length inputs to each of the hardware image processing elements combined with the fixed-point arithmetic implementation requires that the balance between dynamic range and precision be addressed at the design stage. The hardware processing elements are optimised for typical image processing operations; consequently storage and arithmetic operations on long word-length operands (i.e. greater than 16 bits) must be performed by combining operations on sub-word operands, explicitly handling carry and borrow flags, etc.

Synchronous pixel processing at a fixed rate of 20 MHz and hardware support for each of the processing operations discussed in section 3.2 enables a simple check for
viability of the development platform and system throughput to be estimated. The
20 MHz pixel clock rate enables the equivalent of over 75 image data transfers of
spatial resolution 512 x 512 pixels per second; even allowing for the overhead
required to configure processing elements and inter-connections, acquisition and
double-buffering of images from a 60 Hz camera are well within the specification.
Stage 1 of the computation can be completed within one camera frame period using
the ALU elements and the intermediate results stored in VSIMs; each of the NMAC
operations included in stage 2 must be performed sequentially since there is only one
hardware NMAC unit and therefore require two camera frame periods; the arctangent
lookup table operation in stage 3 can be performed concurrently with the second
NMAC operation since no hardware is shared by these operations; stage 4 can be
completed within a single camera frame period by using the available ALU elements.
In conclusion then we can say that it is quite likely that unwrapped phase estimates
corresponding to a single processing frame can be computed before the arrival of the
following processing frame thus meeting the real-time requirement and providing a
throughput of 4×10^6 unwrapped phase sample s⁻¹.

We have confirmed that the generic temporal phase unwrapping algorithm can be
implemented using hardware resources provided by the Datacube pipeline image
processing system, and that the data throughput is both sufficient for real-time
operation and deterministic. It is now possible to look at the design of digital systems
based on this architecture that are tailored for specific whole-field optical
applications. The following chapters describe such systems developed for real-time
profilometry and displacement measurement.

3.6 Summary
The generic digital temporal phase unwrapping system has been presented and data
throughput rates estimated. Implementation choices in the algorithm, hardware, and
software domains have been discussed for the proposed system. Several
implementations have been profiled on the von Neumann architecture, however, data
throughput rates consistently fell below those required for real-time processing of
interferograms acquired at normal video rates. Commercial pipeline image processing
systems were compared and a simple viability study carried out based on a pair of
Real-time digital implementation of the temporal phase unwrapping algorithm

Datacube MV250 processors. Following encouraging results from the study, a
Datacube development platform was purchased in the first quarter of 1996 to enable
design and implementation of real-time profilometry and DSPI systems.

3.7 Tables

<table>
<thead>
<tr>
<th>Number format</th>
<th>Arctangent function implementation details</th>
<th>L1</th>
<th>L2</th>
<th>L3</th>
<th>L4</th>
<th>L5</th>
<th>L6</th>
<th>L7</th>
<th>L8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 FLP</td>
<td>FLP intrinsic function</td>
<td>8</td>
<td>64</td>
<td>64</td>
<td>64</td>
<td>64</td>
<td>64</td>
<td>64</td>
<td>64</td>
</tr>
<tr>
<td>2 FLP</td>
<td>FLP intrinsic function</td>
<td>8</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td>32</td>
</tr>
<tr>
<td>3 FXP</td>
<td>FLP intrinsic function</td>
<td>8</td>
<td>16</td>
<td>16</td>
<td>8</td>
<td>16</td>
<td>16</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>4 FXP</td>
<td>FLP intrinsic function</td>
<td>8</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>5 FXP</td>
<td>FLP heuristic approx</td>
<td>8</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>6 FXP</td>
<td>FXP heuristic approx</td>
<td>8</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>7 FXP</td>
<td>FXP 16x16 LUT</td>
<td>8</td>
<td>16</td>
<td>8</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>8 FXP</td>
<td>FXP 16x16 normalised LUT</td>
<td>8</td>
<td>16</td>
<td>8</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>9 FXP*</td>
<td>FXP 16x16 normalised LUT</td>
<td>8</td>
<td>16</td>
<td>8</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>32</td>
<td></td>
</tr>
</tbody>
</table>

Table 3-1: Details of the nine temporal phase unwrapping algorithms (TPUA1 through TPUA9) profiled on a von Neumann architecture with an Intel Pentium II processor. FLP denotes floating-point number representation. FXP denotes fixed-point number representation. FXP* denotes optimised fixed-point number representation using MMX single instruction multiple data registers. The number format column describes the arithmetic representation used throughout the implementation, with the exception of the arctangent function implementation that is described separately in the following column. L1 through L8 refer to the width of the data paths shown in figure 3-1 expressed as number of bits.
Real-time digital implementation of the temporal phase unwrapping algorithm

<table>
<thead>
<tr>
<th>Datacube MaxTD</th>
<th>Imaging Technology MVC 150 40</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 MHz pipeline processor</td>
<td>40 MHz pipeline processor</td>
</tr>
<tr>
<td>Non-blocking crosspoint (32i/p x 32o/p)</td>
<td>Less flexible crosspoint (16i/p x 14o/p)</td>
</tr>
<tr>
<td>24 Mb image memory per board</td>
<td>4 Mb image memory per board</td>
</tr>
<tr>
<td>C software library</td>
<td>C software libraries</td>
</tr>
<tr>
<td>LynxOs real-time operating system</td>
<td>VxWorks real-time operating system</td>
</tr>
<tr>
<td>Self-hosting Unix environment</td>
<td>Cross development with embedded target</td>
</tr>
<tr>
<td>Protected virtual memory space</td>
<td>Single common address space</td>
</tr>
<tr>
<td>POSIX.1b compliant</td>
<td>Limited POSIX compliance</td>
</tr>
<tr>
<td>Limited 3rd party device drivers</td>
<td>Many 3rd party device drivers</td>
</tr>
</tbody>
</table>

Table 3-2: Comparison of key hardware and software features of the Datacube MV250 and Imaging Technology MVC150/40 pipeline image processing products.
3.8 Figures

Figure 3-1: Schematic representation of a digital system for computing the unwrapped phase change with respect to some initial condition at time $t = 0$. Stage one takes the phase shifted intensity input vector $x(t)$ and computes a pair of images that are proportional to the sine and cosine of the phase change; stage two performs spatial smoothing of these images (as usually required in speckle applications); stage three implements the arctangent function and outputs the wrapped phase change estimate $\Delta \hat{\phi}_w(t,0)$; stage four implements temporal phase unwrapping algorithm based on the modified sum of differences method and presents the unwrapped phase change estimate $\Delta \hat{\phi}(t,0)$ at the output. L1 to L8 denote the binary data word-length at various points in the system.
Real-time digital implementation of the temporal phase unwrapping algorithm

Figure 3-2: Schematic representation of the von Neumann architecture used almost exclusively in most commercial personal computers and workstations. The camera, projector, and display are all interfaced to the general-purpose PCI bus with all data transfers passing through the bridge chipset.
Figure 3-3: Profiling results for each of the nine implementations (TPUA1 through TPUA9) of the digital temporal phase unwrapping system on von Neumann architecture using a 333 MHz Intel Pentium II. The co-ordinate axis shows time taken to complete one processing frame. i.e. 4 phase stepped interferograms with spatial resolution 239 x 192). Each bar graph is divided into four sections corresponding to the processing stages described in section 3.2. The length of each bar representing the mean execution time calculated over 100 processing frames. Errorbars for each section are shown in white. The errorbar for the overall execution time is shown in red.
Figure 3-4: Data throughput figures calculated from the mean execution times shown in figure 3-3. The co-ordinate axis represents the number of number of unwrapped phase values calculated per second. Errorbars are calculated using the minimum and maximum execution times.
Real-time digital implementation of the temporal phase unwrapping algorithm

Figure 3-5: Final wrapped phase difference map $\Delta \phi_w(t,0)$ after 100 processing frames calculated using implementation TPUA1. Black pixels represent $-\pi$ and white pixels $+\pi$ radians. Phase values describe deformation of a thin metal plate subjected to an out-of-plane displacement on the centre of its back surface towards the camera. The final unwrapped phase difference values $\Delta \phi(t,0)$ corresponding to row 112 for all nine von Neumann implementations are shown in figure 3-6. Image dimensions are expressed in camera pixels.
Real-time digital implementation of the temporal phase unwrapping algorithm

Figure 3-6: Unwrapped phase difference values $\Delta \phi(t, 0)$ after 100 processing frames corresponding to row 112 for each of the nine von Neumann implementations. The bottom graph corresponds to implementation TPUA1. Implementations TPUA2 through TPUA9 are shown in sequence and are each offset by 5 radians to prevent the graphs collapsing onto a single line.
Figure 3-7: Computational error introduced in $\Delta \phi(t,0)$ with respect to implementation TPUA1. Figures are calculated along row 112 after 100 processing frames. Standard deviation values are shown in the legend. RMS errors observed using TPUA2 are approximately $-150$ dB with respect to the peak unwrapped phase values. TPUA2* is included to show errors introduced by modifying TPUA2 to simulate an 8-bit data path between the decode block and the complex division. RMS errors observed using TPUA2* are approximately $-70$ dB and are unlikely to be critical for most applications.
Real-time digital implementation of the temporal phase unwrapping algorithm

Figure 3-8: Computational error introduced in $\Delta \phi(t,0)$ with respect to implementation TPUA2*. Figures are calculated along row 112 after 100 processing frames. The graph for TPUA4 has been offset by 0.2 radians.

Figure 3-9: Computational error introduced in $\Delta \phi(t,0)$ with respect to implementation TPUA2*. Figures are calculated along row 112 after 100 processing frames. The graph for TPUA6 has been offset by 0.2 radians.
Real-time digital implementation of the temporal phase unwrapping algorithm

Figure 3-10: Computational error introduced in $\Delta \phi(t,0)$ with respect to implementation TPUA2*. Figures are calculated along row 112 after 100 processing frames. The graphs for TPUA8 and TPUA9 have been offset by 0.1 and 0.2 radians respectively.
Figure 3-11: Schematic representation of commercial pipeline image processing architectures. The CPU dynamically connects sequences of image processing elements (ALUs, LUTs, NMACs, etc.) by configuring a central crosspoint switch and schedules data transfers. The non-blocking nature of the crosspoint switch enables concurrent data transfers between the camera, memory devices, processing element, and displays, unlike the von Neumann architecture.
Real-time digital implementation of the temporal phase unwrapping algorithm

Figure 3-12: Photograph of the Datacube MaxTD pipeline image processing system based on a 5 slot VME chassis. A Motorola MVME167 single board computer is inserted in the left most slot; two MaxVideo MV250 pipeline image processing cards are located in the adjacent slots. An IP format carrier is also present and hosts a general purpose interface bus (GPIB) controller (described in section 6.4.1).
Figure 3-13: Schematic representation of the fixed-point arithmetic unit included on each of the Datacube MV250 image processing boards. The linear and non-linear sections consist of a predetermined binary tree of processing elements that can each be configured separately. The multiple state ALU elements in the non-linear section support a range of functions including addition, subtraction and comparison that can be determined on a pixel by pixel basis depending on the image data value. Both sections can be used concurrently to enable sophisticated image processing operations to be performed efficiently.
Section III: Temporal phase unwrapping algorithm applied to profilometry
4 An instrument for real-time profilometry

4.1 Introduction
The previous chapter discussed issues relating to the implementation of a generic digital temporal phase unwrapping system and the selection of a suitable development platform. This chapter covers the implementation of a high-speed high-precision profilometry system, using the Datacube pipeline image processing development platform discussed in section 3.5, that is able to measure discontinuous test surfaces as easily as continuous ones. The measurement process is based on the principle of spatial frequency scanning of structured light as described in section 1.3.10. The basic idea is to vary the synthetic wavelength of sinusoidal fringes projected on to the test surface over time. For each synthetic wavelength, the fringes are phase stepped according to the standard four-frame phase shifting algorithm described in section 2.2.1 and recorded using a CCD camera. The estimated wrapped phase maps form a three-dimensional phase distribution over the measurement volume. The phase at each pixel is then unwrapped independently along the time axis using the temporal phase unwrapping algorithm (TPUA) and transformed into depth information.

Modifications to the generic TPUA that: (a) increase the unwrapping reliability; (b) increase immunity to phase noise; and (c) reduce the acquisition and computational overheads are discussed. The peripheral components required to implement the system are introduced and the calibration procedure explained. Mapping of the enhanced TPUA to the pipeline image processing hardware is described. The chapter concludes with examples of the profile data obtained using the completed system and examines the measurement accuracy.

4.2 Extending the temporal phase unwrapping algorithm for profilometry

4.2.1 Linear algorithm
Discontinuous surfaces may cause $2\pi$ ambiguity in the recovered phase unless the projected phase range across the surface is limited to a single period. However, the measurement accuracy can be improved by scanning of the synthetic wavelength
An instrument for real-time profilometry during the measurement procedure, increasing in a linear step-wise manner from a single fringe across the field of view at time \( t = 1 \) until a total of \( s \) fringes are projected at time \( t = s \). Note that the phase does not need to be estimated at \( t = 0 \) since this corresponds to the case of zero phase across the entire field of view. The projection optics are arranged so that the central fringe remains fixed in position with additional fringes moving in from both edges of the projection field. All scattering points illuminated by the central fringe will therefore show no change in phase as the number of fringes increases; scattering points at the extreme left and right of the projection field will undergo the greatest phase changes (negative and positive, respectively). Since the phase increment between consecutive phase maps lies in the range \([-\pi, \pi]\) for all scattering points in the projection field the total unwrapped phase change at the time of the \( s^{th} \) phase map measurement (\( \Phi \)) can be calculated simply by summing the wrapped phase differences:

\[
\Phi = \phi(s) - \phi(0) = \sum_{i=1}^{s} \Delta \phi_i(t, t - 1)
\]  

(4-1)

If the error in \( \phi(s) \) is independent of \( s \) then the expected fractional error in the calculated depth values, \( z \), will vary as \( 1/s \) [84].

Zhao et al. [142] noted that the unwrapped phase can be calculated using only the first and last phase maps in the sequence, \( \phi(1) \) and \( \phi(s) \), respectively. Since \( \phi(1) \) automatically lies within the range \([-\pi, \pi]\) we can say that \( \phi(1) = \phi_{w}(1) \) and therefore \( s\phi(1) \) is an approximate estimate for \( \phi(s) \). This estimator can be used to unwrap the final phase map, \( \phi_{w}(s) \), using:

\[
\Phi = U[\phi_{w}(s), s\phi_{w}(1)]
\]  

(4-2)

where we define the unwrapping operator

\[
U(\phi_1, \phi_2) = \phi_1 - 2\pi \text{NINT}\left(\frac{\phi_1 - \phi_2}{2\pi}\right)
\]  

(4-3)

In the case of low-noise data the accuracy of the two methods will be identical, however, Huntley et al. [143] show that the reliability of equation 4-1, that is the probability of recovering the unwrapping phase successfully, is generally more robust. Equation 4-2 amplifies the noise in the low-sensitivity map, \( \phi_{w}(1) \), by a factor \( s \) so
that, when the level of phase noise reaches a standard deviation of $\pi/s$ there will be a significant failure rate. Conversely, equation 4-1 avoids amplification of the noise so that there is only a small probability of failure at any one step; however, all $s$ successive steps must be unwrapped correctly. The probability of failure of the overall unwrapping operation is equal to the probability for a single failed step in the sequence $s$ given by the binomial distribution as $sP$, where $P$ is the probability of failure for a single step. Increasing the number of steps $s$ to attain improved accuracy is therefore also accompanied by a corresponding increase in the probability of failure of the unwrapping procedure that scales linearly with $s$.

Huntley et al.[144] note that the plane on which the scattering point lies is defined by the non-dimensional frequency $\omega$, representing the rate of change of phase with non-dimensional time $t$. An estimate can be calculated as

$$\hat{\omega} = \phi(s)/s,$$

but the accuracy can be improved by fitting a straight-line to the measured phase values in a least-squares sense:

$$\hat{\omega} = \frac{\sum_{i=1}^{s} t \phi(t)}{\sum_{i=1}^{s} i^2}.$$  

(4-5)

Fitting in the least-squares sense using equation 4-5 provides more immunity to noise than simply using the final phase value as in equation 4-4, with root mean square (RMS) error decreasing as $s^{-3/2}$ rather than $s^{-1}$.

### 4.2.2 Forward exponential algorithm

Projecting an exponentially increasing sequence of fringes can significantly reduce the data acquisition and computation time. Although this approach fails to sample the fringes in accordance with the Shannon sampling theorem, the unwrapping is successful because the underlying form of the $\Phi$ versus $t$ graph (a straight-line) is known. An analogy can be drawn with sub-Nyquist interferometry, in which aliased spatial phase maps from optical components with steep slopes can be unwrapped correctly if assumptions are made about the continuity of the first derivative of the underlying height profile[145]. The sequence of measured phase values, for a pixel receiving light from a scattering point to the right of the central fringe (and therefore
increasing phase) is illustrated in figure 4-1(a). $\Delta \phi_w(1,0)$ and $\Delta \phi_w(2,1)$ are both equal to the corresponding unwrapped phase changes, since in each case the number of projected fringes increased by just one, and they can therefore be added to give $\Delta \phi(2,0)$. The wrapped phase map $\Delta \phi_w(4,2)$ will in general contain phase wraps, but can be unwrapped by noting $\Delta \phi(2,0)$ can be used as an estimator for $\Delta \phi(4,2)$. The sum $\Delta \phi(4,2)$ and $\Delta \phi(2,0)$ then results in $\Delta \phi(4,0)$. The unwrapping strategy can be expressed for the general case using recursive equations:

$$\Delta \phi(2t', t') = U[\Delta \phi_w(2t', t'), \Delta \phi(t', 0)] \quad (4-6)$$

$$\Delta \phi(2t', 0) = \Delta \phi(2t', t') + \Delta \phi(t', 0) \quad (4-7)$$

where $t' = 1, 2, 4, 8, \ldots, s$.

The number of measured phase maps required to complete the unwrapping operation using an exponential sequence, rather than a linear sequence as described by equation 4-1, is reduced from $s$ to $(\log_2 s + 1)$ with a corresponding increase in reliability in most situations. Furthermore, unlike equation 4-2 the increasing exponential strategy does not amplify the noise at any stage. The least-squares fitting method can also be used with the exponentially growing fringe sequence, resulting in the estimator

$$\hat{\phi} = \frac{\sum_{s=0}^{\log_2 s} 2^s \phi(2^s)}{\sum_{s=0}^{\log_2 s} 2^{2^s}} \quad (4-8)$$

The root mean square (RMS) error in the phase gradient estimator decreases as $s^{-1}$ and offers only a small reduction in the noise level compared to the basic linear temporal phase unwrapping algorithm.

### 4.2.3 Reversed exponential algorithm

A better approach than the forward exponential sequence from the point of view of reducing measurement error is the reversed exponential sequence illustrated schematically in figure 4-1(b). The algorithm involves starting at the maximum fringe density ($t = s$), and reducing the number of fringes by $1, 2, 4, 8 \ldots s/2$ (i.e. using the values $t = s-1, s-2, s-4, \ldots, s/2$). This ensures that the measured phase values are clustered at the high-$t$ end of the $\Phi$ versus $t$ graph, which improves the least-squares
An instrument for real-time profilometry

estimated gradient when compared to the forward exponential sequence. Unwrapping
is carried out as for the forward exponential method, except that it now proceeds
down from the point \( t = s \) rather than up from \( t = 1 \). The required steps can be
expressed mathematically as follows, starting with the recurrence relations[144]:

\[
\Delta \phi(s - t', s - 2t') = U[\Delta \phi_w(s - t', s - 2t'), \Delta \phi(s, s - t')]
\]

\[
\Delta \phi(s, s - 2t') = \Delta \phi(s - t', s - 2t') + \Delta \phi(s, s - t')
\]

for \( t' = 1, 2, 4, \ldots, s/2 \).

The least-squares estimator for the gradient is now given by:

\[
\hat{\omega} = \frac{s \phi(s) + \sum_{v=0}^{(\log_2 s) - 1} (s - 2^v) \phi(s - 2^v)}{s^2 + \sum_{v=0}^{(\log_2 s) - 1} (s - 2^v)^2}
\]

with the root mean square (RMS) error reduced by a factor of approximately \( \sqrt{\log_2 s} \)
compared with that from the basic linear temporal unwrapping algorithm. In practice
equation 4-11 can be computed more efficiently by maintaining a running weighted
sum of the calculated unwrapped phases:

\[
\hat{\omega} = \Omega(s)\Delta \phi(s, s - 1) + \sum_{v=0}^{(\log_2 s) - 1} \Omega(s - 2^v) \Delta \phi(s - 2^v, s - 2^{v+1})
\]

where the weights \( \Omega(t) \) are given by

\[
\Omega(t) = \frac{s + \sum_{v=0}^{v = t} (s - 2^v)}{s^2 + \sum_{v=0}^{v = t} (s - 2^v)^2}, \quad \text{for } t = s - 2^v (v = 0, 1, 2, \ldots, (\log_2 s) - 1),
\]

and

\[
\Omega(s) = \frac{s}{s^2 + \sum_{v=0}^{(\log_2 s) - 1} (s - 2^v)^2}.
\]

Pre-calculating the weights and storing them in a lookup table can reduce
computational overhead.

The improved phase unwrapping reliability attained by reducing the number of phase
maps required, improved immunity to noise, and efficient computation steps makes
the reversed exponential method an attractive solution for real-time high-precision profilometry.

### 4.2.4 Fourier Transform Ranging Method

An alternative method for analysing the data that avoids the need for phase unwrapping is introduced in reference [144]. This technique is based on the Fourier analysis of the two-dimensional intensity distribution (i.e. along the time, \( t \), and phase-shift index, \( p \), directions) measured at each pixel. The transform consists of a DC peak with two sidelobes; if the phase-shifting algorithm used covers one complete cycle then the sidelobes will always appear on the lines \( k_p = \pm 1 \). In practice, evaluation of the full 2-D transform \( H(k_p, k_i) \) is unnecessary, and the discrete Fourier Transform is only evaluated at the point \( k_p = 1 \):

\[
H(k_p, k_i) = \sum_{p=1}^{q} \sum_{t=1}^{q} I(p,t) \exp(-i2\pi \left( \frac{k_p(p-1)}{q} + \frac{k_i(t-1)}{s} \right))
\]

Computationally the aim is to find the value \( k_i = \kappa \) that maximises \( |H(1,k_i)|^2 \); the corresponding estimator for \( \omega \) is then given by:

\[
\hat{\omega} = \frac{2\pi\kappa}{s}
\]

Whilst currently unsuitable for real-time profilometry due to the high computational overhead and data dependent run-times, Fourier Transform ranging does offer two principle advantages: (a) improved reliability with increasing fringe density \( s \) (c.f. decreasing reliability of unwrapping methods with increasing \( s \)); and (b) improved tolerance to multiple reflections (which can result in multiple isolated peaks in the 2-D Fourier domain).

### 4.3 Peripheral hardware components

#### 4.3.1 Fringe projection system

Both coherent and incoherent projected structured light patterns may be used to implement profilometry systems, as discussed in section 1.3.10. Coherent projection systems described in the literature typically generate a fringe field using two coherent fibre optic point sources separated by a small distance. Fine adjustment of the fibre
separation using a mechanical translation device enables the fringe pitch to be altered, however, the process is relatively slow and is unsuitable for projecting rapidly changing fringe patterns. Similarly, incoherent projection systems have been described that use a conventional slide projector together with a sequence of pre-prepared gratings, however, these low-cost systems also have relatively slow update rates.

There are two dominant white light spatial light modulator (SLM) technologies that may be considered for high-speed fringe projection: (a) liquid crystal display (LCD) technology; and (b) digital mirror device (DMD) technology. LCDs act as a light shutter that modulates the amount of polarised light that can be transmitted through the panel. Liquid crystal is sandwiched between two orthogonal polarisers and adjusts the light polarisation according to the applied analogue signal. Thin film transistor (TFT) LCDs require a transistor to control the polarization at each pixel on the LCD panel and are manufactured using either amorphous silicon (am-Si) and polycrystalline silicon (poly-Si). Am-Si LCDs are built by depositing transistors on a large glass substrate with a transistor located at the corner of each pixel. These single panel devices are available in monochromatic and multi-colour formats but typically achieve only low pixel packing densities (so-called fill factor) and relatively poor image quality. Poly-Si LCDs are fabricated at high temperatures on quartz substrates and are much smaller than am-Si panels. These devices are exclusively monochromatic and use smaller transistors enabling much higher fill factors and superior image quality.

Projector displays based on DMD are inherently digital in nature. The DMD is a thumbnail-size binary spatial light modulator developed by Texas Instruments and consists of an array of movable micro-mirrors mounted over a CMOS static random access memory (SRAM). Each mirror is used to modulate reflected light and is independently controlled by loading data into the memory cell located below the mirror. The data electrostatically controls the mirror’s tilt angle in a binary fashion, where mirror states are either +10 degrees (ON) or -10 degrees (OFF). Light reflected by the ON mirrors is then passed through a projection lens and onto a screen. The mirror states can be switched several thousand times every second and enable greyscale images to be generated simply by controlling the ratio of ON and OFF
An instrument for real-time profilometry

states at each pixel during a single frame display period. Since DMD technology is based on reflection the control electronics can be positioned below the mirror elements allowing a higher fill factor than can be obtained using LCD technology and reducing the pixelisation effect generated by LCD projectors. The reflective DMD technology is more efficient than LCD technology at getting light from the lamp to the screen, with future increases in pixel resolution expected to bring corresponding improvements in efficiency. Furthermore, fast mirror-state switching enables truly independent images to be projected at frame rates much higher than can be achieved using LCD technology that typically suffers from hysteresis effects.

A Proxima DP4200 colour data projector (shown in figure 4-2) was chosen that utilises a Texas Instruments DMD spatial light modulator (800 x 600 pixels) capable of updating the projected scene at 60 frames s\(^{-1}\). The device uses a colour wheel to generate colour images, however, the wheel can easily be removed to increase the projected light output by a factor of around three for greyscale images to 1400 ANSI lumens. The high frequency time-division switching of the Proxima DMD array was recorded by placing a small photodiode (AME SPOT-4DMI) in the image plane such that light falling onto the sensor corresponded to a single pixel mirror element. The potential difference across the photodiode was amplified using a TL082 operational amplifier and recorded using a Gould 1602 digital oscilloscope, together with the vertical synchronisation component (Vsync) of the analogue SVGA video signal driving the projector. Three independent snapshots were recorded at each of the projection intensities 0, 32, 64, 96, 128, 160, 192, 224, 255. Figure 4-3 illustrates typical examples of the trace snapshots recorded, corresponding to projected scenes with greyscale pixel intensities 0, 128, and 255. The photodiode signal corresponding to a single video frame lies between adjacent falling edges in Vsync signal. The photodiode signals were normalised into the range [0,1] and the mean intensity over a single frame period calculated for each snapshot recorded. Figure 4-4 shows the mean normalised photodiode signal for each of the sampled projector pixel intensities. The minimum projection distance (1.5 m) was reduced by mounting a 1 diopter close-up lens (72 mm diameter) in front of the built-in projector lens using a custom machined lens bezel. This enabled the projector to be focused on a plane around 0.8 m from the lens giving an illuminated field of around 0.25 m x 0.25 m.
4.3.2 Digital CCD camera

Unwrapping a 3-D phase data set along the time axis requires that each of the constituent 2-D phase maps are geometrically aligned. Consequently, care must be taken to ensure that the geometry of the pixel data acquired using a CCD camera is maintained and not corrupted by the data transmission or digital to analogue conversion process.

CCD cameras based on analogue data transfer typically use an industry standard composite video signal (e.g. CCIR or RS170) that encodes the vertical and horizontal timing information together with the image data. The receiver (often referred to as the frame-grabber) must then decode the appropriate synchronisation from the composite video input. The vertical synchronisation signals (Vsync pulses) are detected using a combination of a low pass filter and an edge detector; the horizontal signals (Hsync pulses) are detected using a phase locked loop (PLL) circuit. The clock frequency generated by the PLL will fluctuate for some time after the image capture until a stable lock is obtained. Imprecision in the phase comparison will lead to the phenomenon known as line jitter that causes horizontal displacement of lines within the image. Direct transfer of the Hsync and Vsync signals can provide for better synchronisation stability than composite video or composite synchronisation signals. However, pixel-synchronous sampling must be used, rather than a PLL circuit, to eliminate line jitter. Hsync and pixel clock signals supplied by the camera enable the frame-grabber to eliminate sub-pixel jitter since the Hsync pulse can be isolated to a single pixel clock pulse. Jitter by whole pixels is possible, however, in general pixel-synchronous sampling results in a stable geometry, elimination of line jitter, and one-to-one mapping of sensor elements to image pixels. The notable disadvantage of pixel-synchronous sampling is that the high-frequency pixel clock limits the signal transmission range.

CCD cameras based on digital data transfer are available which use a variety of industry standard and proprietary protocols. Virtually noise-free image transfer between the camera and the image processing system can be achieved by using digital transmission protocols with pixel-synchronous sampling.
A 60 Hz EEV CAM17 synchronous camera with a 512 x 512 pixel CCD array was chosen to acquire the fringe patterns. The camera has an 8-bit RS422 digital output for image data and independent pixel clock, Hsync, and Vsync outputs. The CCD provides the master pixel clock and framing pulses; the pipeline processor is then configured as a slave to this external clock to ensure that the video signal applied to the projector is synchronised to the camera. Transmitting the digital image data, and more importantly the pixel clock and framing pulses, over long cables between the camera and the MV250 introduces a small time delay. By designating the camera as the source of the timing signals we ensure that all of these channels are delayed by equal amounts and remain in phase.

Shortly after installation of the camera by Optimum Vision it became clear that the poor quality digital cabling and plastic connectors used to connect the camera to the image processor were susceptible to electrical interference and causing whole frames to be lost. Further investigation revealed that the EEV camera specification was only capable of transmitting the 20 MHz image data over cable lengths of less than 1 m, severely limiting the deployment flexibility of the measurement system. A module was therefore designed which receives the 8-bit RS422 digital image data and timing signals from the camera and includes an electronic circuit to drive the signals over longer cable lengths. The circuit schematic and PCB layout are included in appendix A6. In addition, 5 m of industrial grade multi-core shielded cable and metallic connectors were fitted to complete the transmission system. The drive module can clearly be seen mounted directly behind the camera in figure 4-2, together with the Nikon standard 35 mm camera lens (focal length 50 mm) used predominantly in the following chapters.

### 4.3.3 Support structure

The support structure for the shape measurement apparatus provides a common stable platform for the camera and projector, and another platform for the translation stage. Three magnetic bases are used to clamp each platform securely to an optical table ensuring that the relative position of the equipment is constant during experiments.
The camera and projector platform consists of a crossbeam along which the camera and projector mounts can slide to allow the angle between the optical axes to be adjusted. Since the projector has been manufactured such that its optical axis is elevated above the horizontal (suitable for presentation applications) the projector mount has been inclined at an angle to compensate and provide a horizontal optical axis. The camera mount provides three axes of movement: the height and elevation can be adjusted so that the camera and projector optical axes are vertically aligned; and the azimuth angle can be adjusted to align the camera with the measurement volume.

The supplied projector enclosure and support legs are machined from plastic and therefore compromise the stability of the projector during the severe temperature changes experienced during experiments. The projector has therefore been modified to provide fixing bolts to attach the projector to its mounts on the support platform. The fixing bolts clamp the projector’s principal metal chassis and optical subassembly directly to the supports. Spacing holes through the plastic enclosure ensure that variations in the dimensions of the plastic do not affect the position of the optical subassembly, and hence the optical axis of the projector.

### 4.3.4 Translation stage

An Aerotech ATS 15030 translation stage provides a stable base on which objects can be translated through the measurement volume. The stage is positioned using a 0.2 inch pitch lead-screw driven by a stepper-motor and whilst the quantised rotation steps cannot match the resolution provided by servo-drives, the stepper-motor approach ensures that the stage remains stationary for constant drive signal; servo-drives on the other hand tend to introduce jitter due to variations in the feedback circuitry. The translation resolution of the stage is 1/10000 inch with repeatability of 1/2500 inch. Straightness and flatness of travel is specified as 2 μm per 25 mm of travel. The stage has been modified with mounts for a right-angled plate that can be used during the calibration procedure and mounts for various test surfaces. A goniometer has been mounted on a rotation stage so that the orientation and position of test samples can be precisely controlled. The translation stage has sufficient travel (300 mm) such that the right-angled plate can be used while calibrating the
measurement volume, then when no longer required, translated out of the way. The mounts for test surfaces on the other end of the stage are then conveniently located within the measurement volume.

4.4 Calculating surface profile from phase data

The camera pixel coordinates \((m,n)\) uniquely define a straight-line passing through both the pixel and the centre of the camera lens. The line intersects the planes of constant phase gradient generated by the projected fringe sequence. A phase gradient estimator \(\hat{\phi}\) measured at pixel \((m,n)\) therefore uniquely defines the \((X,Y,Z)\) coordinates of the scattering point on the test surface. Modelling the \((\hat{\phi},m,n)\) to \((X,Y,Z)\) transformation is in principle straightforward but complicated by the number of degrees of freedom involved in setting up the camera and projector. The approach used throughout this thesis therefore involves calibration of the system following alignment of the camera and projector as described by Saldner et al. [146].

A glass flat is translated through the measurement volume in the \(z\)-direction and the phase gradient estimator map \(\hat{\phi}(m,n)\) recorded at approximately 20 sample positions along the axis. The quadratic function given in equation 4-17 is then fitted to the data using least-squares minimisation.

\[
Z(m,n) = A_2(m,n)\hat{\phi}(m,n)^2 + A_1(m,n)\hat{\phi}(m,n) + A_0(m,n) \quad (4-17)
\]

A calibration mask image, \(\mu_c\), is generated to mark those pixels that have low fringe modulation or at which the least-squares fitting did not converge; the image can subsequently be used to remove pixels at which we have low-confidence in the calibration data set.

The demagnification coefficient (i.e. object size/image size) is approximated by attaching horizontal and vertical rules to the glass flat and translating through the measurement volume in the \(z\)-direction. Images are recorded at two or more positions with the \(x\)- and \(y\)-components of the demagnification coefficient calculated as:

\[
\Gamma_x(Z) = \frac{X - X_0}{m - m_0}, \quad \text{and}
\]

\[
\Gamma_y(Z) = \frac{Y - Y_0}{n - n_0}, \quad \text{respectively,}
\]

121
where \((m_0, n_0)\) is the pixel at the centre of magnification. The centre of magnification \((X_0, Y_0)\) can be taken as \((0,0)\) if the measurement volume coordinate system is arranged such that the z-axis lies along the camera axis. Since the EEV camera has a square pixel structure the data analysis is simplified by using a global demagnification function, \(\Gamma(Z)\), calculated as the average of the \(x\)- and \(y\)-components. Finally, the demagnification function can be interpolated by fitting to a quadratic function in a least-squares sense.

The calibration data enables the co-ordinates to be computed using a two step process: (a) the phase gradient map \(\hat{\omega}(m, n)\) is converted to depth map \(Z(m, n)\) using the quadratic approximation given in equation 4-17; and (b) the \(X(m, n)\) and \(Y(m, n)\) maps are calculated from the demagnification coefficient as:

\[
X(m, n) = \Gamma[Z(m, n)](m - m_0), \quad \text{and} \quad (4-20)
\]
\[
Y(m, n) = \Gamma[Z(m, n)](n - n_0). \quad (4-21)
\]

4.5 Algorithm implementation on the pipeline image processing system

4.5.1 Introduction

Pipeline image processing essentially converts an image consisting of a 2-D array of pixel values into a 1-D serial stream of individual pixels. The pixel stream can be processed one pixel at a time by a series of specialised processing elements, and then transformed back into a 2-D image. At the beginning and end of each 1-D pipeline is a 2-D image surface that is stored on a virtual surface image memory (VSIM) hardware device. Pipeline processing has an important performance advantage over fixed-data-path, general-purpose image processing architectures, which process whole images as discrete frames. This advantage is derived from placing each specialised hardware processing element at the particular point in the 1-D stream where it can perform its individual processing task on each pixel, without waiting for the entire frame to be processed.

The processing hardware can be assembled in a variety of configurations that must be reorganised very quickly to implement a pre-compiled sequence of computation tasks.
Pipe-operation altering threads (PATs) are used to reduce the time required to perform these reorganisation tasks such that they can be completed during the camera inter-frame period. The host CPU calculates the hardware register values required to construct the processing pipeline configuration and set up the delay elements to temporally align the data operands. These calculations can typically take several seconds to complete, whilst the time required to write the results to the hardware registers is only of the order of tens of microseconds. Where pipelines are reused, therefore, a pre-calculated PAT is defined that contains only the essential hardware instructions and register loads. The PAT resides in the operating system kernel so that PAT-related events occur at a higher priority than the user-space application program ensuring that reorganisation of the pipeline hardware is not delayed by other system tasks. In addition, a control PAT is defined that schedules the other PATs by passing parameters and semaphores between them to reduce system latency. Embedding control in the operating system kernel in this way avoids the delay associated with context switching between the kernel and user-space applications that can be comparable to, and sometimes longer than, the time required to load the hardware registers (e.g. a delay of approximately 60 µs when running a PAT from user space).

The following paragraphs describe the image processing pipelines that are assembled by these PATs and the sequence in which they are applied.

Simplified high-level schematics of the processing pipelines are included in appendix A7 labelled S1 through S7 and can be regarded of a snapshot of the hardware at a given point in time. The complete set of image processing hardware in use at any one time cannot exceed that which is available on the pair of MV250 cards described in section 3.5. Connections between hardware elements are represented by interconnecting lines with the direction of data flow represented by arrows; the width of the data path (expressed in bits) is shown alongside. Multi-byte data values must be stored across several image surfaces since the VSIM hardware devices cannot store word lengths longer than 8 bits. The subscripts B0, B1, ..., etc. are used to denote byte number in a multi-byte data value; B0 represent the least significant byte, B1 the next most significant byte, etc. The digital camera is represented schematically by a virtual image surface denoted CamB0 that acts as a source for an image data stream. The analogue video signal generators driving the projector and monitor displays are
schematically represented by virtual image surfaces denoted DAC0 and DAC1, respectively, and act as destinations for image data streams. Two pipelines continually stream data to these virtual surfaces to ensure that the displays are constantly refreshed. It can be seen that at any time there are several pipelines streaming data in parallel. For example, hardware arrangement S1 includes a data stream from the camera (CamB0) to the surface I01B0; a data stream from the surface FringeB0 to the surface ProjB0; a data stream from the surface ProjB0 to the surface DAC0; a data stream from surfaces DispB1B0, and OvlyB0 to surface DAC1; and a complex data stream that terminates on surfaces φB1B0 and ΔφB1B0.

The hardware arrangements S1 through S4 are applied in sequence and correspond to the projection and acquisition of the four phase stepped fringe patterns. The user is able to select the maximum fringe density, \( s \), used during the measurement process from a list of supported values (64, 32, 16, 8 and 4). The sequence of hardware arrangements is applied once for each of the fringe densities in the reversed exponential series (e.g. when \( s = 32 \), the sequence S1, S2, S3, S4 is applied once for each of the fringe densities 32, 31, 30, 28, 24, 16). One extra sequence is performed at the end corresponding to zero phase across the projection field in which the phase map is simulated rather than measured to reduce errors. A uniform white scene is projected during this sequence and enables the test surface texture to be measured with no additional time overhead. Finally, the hardware arrangements S5 through S7 are applied in sequence to transform the phase gradient estimator to depth and update the monitor display. The total measurement process using \( s = 32 \) fringes takes 33 camera frame periods (i.e. 0.55 s); 2 frame periods to preload the projector and camera memory buffers, 7 cycles of the sequence S1, S2, S3, S4 to compute the phase gradient estimator (i.e. 28 frame periods), and 3 frame periods to apply S5, S6 and S7.

4.5.2 Fringe projection

During the program initialisation period a lookup table is created that contains the projected fringe pattern data. The table is compressed to conserve image memory using the property that each line in a given projected fringe pattern is identical. It is therefore only necessary to store the first line of each pattern: a reversed exponential sequence of phase-stepped images with \( s = 16 \) fringes (i.e. \( t = 16, 15, 14, 12, 8 \)), for
An instrument for real-time profilometry

example, requires a table with 20 rows of pixel data occupying approximately 20 kb of memory as shown in figure 4-5. The lookup table data is stored in image surface \( \text{Fringe}_0 \) and expanded at run-time using a dedicated pipeline to reconstruct the complete projected image that is written to surface \( \text{Proj}_0 \). Surface \( \text{Proj}_0 \) is double-buffered and acts as a source for the continuous display pipeline that feeds the analogue video signal generator (DAC0) that drives the projector.

### 4.5.3 Image acquisition

A dedicated acquisition pipeline handles 8-bit digital image data received from the CCD camera and computes the \( N(t) \) and \( D(t) \) images (which are proportional to the sine and cosine of the wrapped phase, respectively) according to equation 2-52. The acquisition pipeline performs the subtraction operation using a simple 8-bit arithmetic element located on the VSIM hardware device leaving the more powerful arithmetic units (AU0 and AU1) available for use by the main computation pipeline that runs concurrently. The 8-bit data-path through this simple arithmetic element requires that the least-significant bit of the 9-bit subtraction result is discarded and is equivalent to the 8-bit subtraction investigated using the von Neumann architecture in section 3.4.

In hardware configurations S1 and S2 the acquisition pipeline reads the image data stream \( I(0) \) and \( I(1) \), respectively, from the camera and writes it directly to the image surface \( I01_0 \). The surface receive gateway element ensures that the individual images \( I(0) \) and \( I(1) \) are stored in separate regions of the image memory and are not overwritten. In hardware configuration S3, the image data received from the camera, \( I(2) \), is subtracted from the image stream read from surface \( I01_0 \); the surface transmit gateway ensures that the region of memory selected for reading corresponds to \( I(0) \). The resulting image stream is written to surfaces \( N_0 \) and \( D_0 \); the receive gateway for \( N_0 \) selects a region of image memory that is reserved as a scratch buffer and is never read, and the receive gateway for \( D_0 \) selects a region of memory reserved for \( D(t) \). In hardware configuration S4, the receive gateway for surface \( I01_0 \) selects the region of memory corresponding to \( I(1) \) for reading and the resulting data stream is subtracted from the data stream received from the camera, \( I(3) \). The subtraction result is written to destination surfaces \( N_0 \) and \( D_0 \); this time,
however, the receive gateway for $N_{B0}$ selects an active region of memory and the receive gateway for $D_{B0}$ selects the region of memory reserved for the scratch buffer.

Manipulating the regions of memory selected by the surface receive and transmit gateways in this way appears at first to complicate the system design, however, it enables efficient switching of data paths. Gateway settings can be modified with little or no overhead whereas modifying the underlying pipeline structure requires significant computational effort. The acquisition pipeline has therefore been implemented using two basic configurations (i.e. $S_1$ and $S_2$ share the same basic configuration, as do $S_3$ and $S_4$). The modified region of interest approach is used extensively throughout the pipeline processor implementation and facilitates double buffering of the image data.

4.5.4 Mapping reversed exponential algorithm to hardware

Hardware configuration $S_1$ includes the acquisition, projection and display pipelines discussed above, however, the majority of the hardware processing elements are used to compute the estimated wrapped phase map stored in image surfaces $\phi_{W_{B1:B0}}$ and the unwrapped phase difference stored in image surfaces $\Delta \phi_{B1:B0}$. The wrapped phase is computed using a simple four-quadrant arctangent lookup table that takes the 8-bit $N(t)$ and $D(t)$ images (which are proportional to the sine and cosine of the wrapped phase, respectively) as operands. The operands are not normalised as investigated in section 3.4 as the acquired white-light interferograms are generally well modulated; normalisation becomes more important when dealing with speckle interferograms in which a large number of the pixels are poorly modulated. The lookup table result lies in the range $[-\pi, \pi]$ and is represented using a signed 16-bit fixed-point data format. The $\phi_{W_{B1:B0}}$ VSIMs are double buffered to allow the current wrapped phase map to be written to memory whilst the previous wrapped phase map in the reversed exponential sequence is being read from the same device. Constant $K_1$ is a mask applied to the previous phase map data stream using the logical AND operator and is normally set to all-ones (i.e. all bits are set to 1) to allow the data to pass unmodified. The non-linear section of the arithmetic unit AU0 (denoted AU_N0) computes the wrapped phase difference $\Delta \phi_w (s-t', s-2t')$ for $t'=1,2,4,\ldots,s/2$. The linear arithmetic sections
An instrument for real-time profilometry

\(\mathrm{AU}_0\) and \(\mathrm{AU}_1\) unwrap the phase difference according to equation 4-9 using the running sum stored on surfaces \(\Sigma \Delta \phi_{B1:B0}\). The coefficients \(K_2\) and \(K_3\) are used to scale the fixed-point operands to the subtraction operator such that their bit-fields are aligned (i.e. both operands have the same fixed-point scaling). The unwrapping operator expressed in equation 4-3 is implemented using a shift operator to perform the division by \(2\pi\); a least-significant bit rounding element performs the NINT operation. Constant coefficient \(K_4\) and \(K_5\) are used to scale the fixed-point operands to the subtraction operator such that their bit-fields are aligned. Constant mask \(K_6\) is set to zero on the first application of hardware configuration S1 to reset the image surfaces \(\Delta \phi_{B1:B0}\); on subsequent applications the mask is set to all-ones to allow the unwrapped phase data to pass unmodified.

Hardware configuration S2 uses the unwrapped phase difference computed by S1 to update the phase gradient estimator map, \(\tilde{\phi}\), stored in image surface \(\omega_{B1:B0}\) and the running sum of unwrapped phase values stored in image surfaces \(\Sigma \Delta \phi_{B1:B0}\). Each of these destination surfaces is double buffered to enable the previous values to be read whilst concurrently writing the new values to the same device. Each term in equation 4-12 is calculated by multiplying the unwrapped phase difference data stream by the weighting factor specified in equations 4-13 and 4-14 using constant coefficients \(K_7\) and \(K_8\). The precision of the fixed-point multiplication is enhanced by multiplying the weighting factors by a constant scaling factor, \(\Omega_0\), such that the modified weightings can be represented by integer values. The scaling factor depends on the maximum fringe density and is numerically equal to the denominator in equations 4-13 and 4-14:

\[
\Omega_0 = s^2 + \sum_{r'=0}^{(\log_2 s)^{-1}} (s - 2^r)^2
\]  

(4-22)

Hardware configuration S3 updates the minimum fringe modulation (or more precisely, the modulation squared) observed at each pixel during the measurement process. Modulation is estimated from the \(N(t)\) and \(D(t)\) images (proportional to the sine and cosine of the wrapped phase, respectively) and compared to the previous minima stored in surface \(m_{2B0}\).
An instrument for real-time profilometry

Hardware configuration S4 includes the usual acquisition, projection and display pipelines but does not perform any additional computation.

The number of cycles through the sequence of hardware configurations S1, S2, S3, S4 is determined by the maximum fringe density, \( s \). Once these cycles have been completed hardware configuration S5 scales the fixed-format phase gradient estimator data to correct for the weighting scale factor, \( \Omega_0 \), applied in S2 such that the data now lies in the range \([ -\pi, \pi ]\). Subsequent processing of the estimator data is no longer dependent on the maximum fringe density used during the measurement process. Image data acquired whilst projecting a uniform white scene is written to the image surface \( \text{Tex}_{B0} \) and provides a record of the test surface texture.

Hardware configuration S6 performs the conversion from phase gradient estimator map, \( \hat{\omega}(m,n) \), to depth map, \( Z(m,n) \), following the quadratic approximation expressed in equation 4-17. Image surfaces \( A2_{B1:B0}, A1_{B1:B0}, \) and \( A0_{B1:B0} \) hold the coefficients determined during the calibration procedure described in section 4.4. Pre-calculated \( \hat{\omega}^2 \) values are stored in the AP_WORD lookup table since all the hardware multiply elements are employed elsewhere in the pipeline. The resulting depth map is written to image surfaces \( z_{B1:B0} \). The second calibration step of converting the \((m,n,Z)\) co-ordinates to true \((X,Y,Z)\) co-ordinates is not implemented on this system, but could be achieved at the cost of an additional frame period.

The final hardware configuration, S7, is responsible for updating the image surfaces which feed the monitor display. Constant coefficients \( K10 \) and \( K11 \) are used to offset and scale the depth information before it is used as an index to the colormap stored in the AP_WORD lookup table. The non-linear section of arithmetic unit AU1 detects pixels that should be masked from the result either due to poor fringe modulation or saturation of the CCD elements. Poorly modulated pixels are detected by thresholding the modulation data maintained in image surface \( m_{B0} \); pixels at which the CCD element has saturated are detected by thresholding the texture map stored in image surface \( \text{Tex}_{B0} \). This information is combined with the mask recorded during the calibration procedure to generate the overlay image, which is written to image surface \( \text{Ovly}_{B0} \). The overlay image is subsequently used in the display pipeline to mask depth results on a pixel-by-pixel basis.
4.5.5 Calibration of SLM / CCD subsystem

Non-linearities in the projector spatial light modulator (SLM) and camera CCD result in distortion of the sinusoidal intensity profile in the acquired images. An example of the acquired intensity profile generated from unmodified sinusoidal fringes (figure 4-6) is shown in figure 4-7. Global intensity variations are also apparent that are thought to be due to a combination of initial non-uniformity, use of a simple close-up lens to reduce the projection field, and a reference flat that had a significant specular component in the scattered light. These variations could have been reduced by the use of a compensating variation of the transmission of the SLM[147]. The non-linear effects are, however, more of a problem since they can result in significant systematic phase errors at twice the fringe frequency[102].

The transfer function of the combined SLM / CCD system was characterised by projecting a sequence of uniform grey images, with increasing intensity over time, and plotting the acquired intensity against grey level written to the projector. The inverse transfer function was applied to the sinusoidal projected fringes to produce a modified projection pattern (figure 4-6). Applying the discrete Fourier transform (DFT) to the zero-padded acquired intensity profile (figure 4-8) revealed that the second harmonic had been significantly reduced. Higher harmonics (in particular the third harmonic to which the four frame phase stepping algorithm is sensitive) do not appear to be a significant problem.

4.6 Application of the real-time profilometer

The technique of shape measurement by projected fringes and temporal phase unwrapping was demonstrated in a number of papers[143,144,146] and one of these also validated the performance of the reversed exponential method using floating-point arithmetic to calculate the depth maps. This section therefore aims to show that results obtained by the pipeline processor are in close agreement with those obtained using floating-point arithmetic, rather than to re-validate the basic technique. This was done by measuring the profile of a flat glass plate spray-coated with matt white emulsion paint at a sequence of arbitrary Z values. The angle between the optical axes of the projector and the camera was 35°. The depth values were calculated using the image pipeline processor implementation of the reversed exponential algorithm in...
An instrument for real-time profilometry.

An instrument for real-time profilometry. Regions of interest (ROIs) from the acquired images were stored in the unused virtual surface image memory for post processing. The mathematical software package Matlab (developed by The Mathworks Inc.) was used to analyse these acquired images offline and compute the depth profile using floating-point arithmetic. Figure 4-9 shows one cross-section of a plane at positioned at \( Z = -25.4 \) mm, which illustrates a typical result. The mean depth calculated by the pipeline processor along the cross-section is \(-25.4518 \) mm, with a standard deviation of \( 31.1 \) µm which is \( 1/4823 \) of the width of the measurement volume (150 mm). The difference between the depth profile calculated by the pipeline processor and Matlab from the same set of acquired images is shown in figure 4-10. The mean error along the cross section is \(-10.2 \) µm with a standard deviation of \( 17.3 \) µm. The largest contributory error source is likely to be the loss of the least-significant bit of the 9-bit subtraction results \( I_0 - I_2 \) and \( I_3 - I_1 \) in stage one of the TPUA as suggested during discussion of the algorithm implementation in section 3.4. To retain accuracy, this rounding step was not implemented in the floating-point evaluation. The mean depth calculated over all unmasked pixels in the depth map (rather than along a single line) computed by the pipeline processor is \(-25.4433 \) mm with a standard deviation of \( 34.0 \) µm which is \( 1/4335 \) of the width of the measurement volume.

Finally, an example application of the profiling system to a practical engineering component is presented. Figure 4-11 shows a photograph of a machined die-cast aluminium pump housing; the component has a wide range of surface finishes, including rough die-cast surfaces in the recesses and smooth specular regions that have been milled during the machining process for locating O-ring seals. No additional surface treatment was applied to the component during the measurement process. Figure 4-12 shows the wrapped phase map from the first cycle of projected fringe patterns \( (s = 32 \) fringes) and contains many singularities introduced by irregular physical discontinuities in the test surface and shadowing effects: placement of boundaries to spatial unwrapping is extremely difficult, however, unwrapping along the time axis avoids these problems. Figure 4-13 shows the result of using the pipeline image processor to unwrap using the reversed exponential temporal phase unwrapping algorithm followed by conversion to the \( (m,n,Z) \) domain. The result was...
displayed as a pseudo-colour image on the system monitor 33 camera frame periods (i.e. 0.55 s) after the start of the measurement process; masked pixels are denoted as dark blue regions. Figure 4-14 shows a 3-D mesh representation of the same data set; masked pixels have been omitted from the plot and the mesh has been sub-sampled so that only 1 in 5 of the available samples along each axis are represented (i.e. only 1/25 of the data points are plotted).

4.7 Summary

In general terms, it has been shown that pipeline image processing technology can be used to perform real-time quantitative analysis, including real-time phase unwrapping of phase-shifted fringe patterns acquired at normal video rates (60 frames s⁻¹). When applied to the measurement of surface profiles using the reversed exponential algorithm, 24 fringe patterns of varying pitch and phase are projected and analysed, enabling approximately 250,000 co-ordinates to be measured and displayed as a pseudo-colour image on the computer monitor 0.55 s after the start of the measurement process. Low fringe contrast pixels, and pixels at which the sensor has saturated, are automatically removed using an overlay mask. Data for a given pixel are analysed independently of the other pixels in the image so that objects with discontinuous surfaces are profiled as easily as continuous ones. Depth measurement accuracy is better than 1/4000 of the measurement volume dimension using a maximum fringe density of s = 32 fringes. Errors introduced by the fixed-point arithmetic on the pipeline processor are not significant compared to other sources of error (e.g. optical and electrical) in the system when using a maximum fringe density of s = 32 fringes.
Figure 4.1: Unwrapped phase for a given pixel measuring the light scattered by a point to the left of the central fringe: (a) forward exponential method; (b) reversed exponential method. Here \( t \) is the total number of fringes; \( s \) is the maximum value of \( t \). The numbers above the circles indicate the order in which the phase values are acquired.
Figure 4-2: Photos of the peripheral systems used in the high-speed profilometry system. The top photo shows the Proxima DP4200 digital mirror device (DMD) projector and EEV CAM17 camera mounted on a steel support rig. The machined aluminium bezel used to mount a close-up lens in front of the projector can be seen. The EEV camera is fitted with a Nikon standard 35 mm lens with focal length of 50 mm. The camera is unable to drive digital image data over cable lengths greater than 1m. The lower-left photo clearly shows the custom made RS422 digital line driver mounted directly behind the camera that amplifies the signals enabling a 5m cable to be used without signal loss. The lower-right photo shows the Aerotech electronic translation stage used in the calibration procedure. A right-angled plate at the back of the stage is used to mount the reference plane during calibration and a goniometer at the front is used to position test surfaces.
Figure 4-3: Oscilloscope traces to illustrate the time-division switching of a single digital mirror device (DMD) element. Channel 1 corresponds to the vertical synchronisation signal (Vsync) applied to the Proxima DP4200 projector. Channel 2 corresponds to the amplified voltage across a single photodiode positioned in the projected field of view. Traces (a), (b), and (c) were generated using a projected greyscale intensity of 0, 128, and 255 respectively.
Figure 4-4: Mean normalised photodiode signal intensity for each of the sampled projector pixel intensities. Errorbars calculated using a sample size of three independent random snapshots of a single video frame.

Figure 4-5: Compressed lookup table used to store vertical projected fringe pattern data; only the first line of each pattern is stored to conserve memory. Each row of the LUT is expanded at run-time using an image pipeline to create the complete projected fringe pattern as illustrated.
Figure 4-6: Intensity profile of projected fringes. Unmodified sinusoidal profile and modified profile with inverse SLM / CCD transfer function applied as shown.

Figure 4-7: Intensity cross section \( I(m, n_0) \) of acquired image (row \( n_0 = 250 \)) using the projected fringes shown in figure 4-6.

Figure 4-8: Discrete Fourier transform (DFT) \( \tilde{I}(k_m, n_0) \) (with linear vertical scale) of the acquired image intensity profiles shown in figure 4-7. Units for spatial frequency \( k_m \) are the number of cycles across field of view. The small periodic ripple is a consequence of the zero-padding and rectangular data window.
Figure 4-9: Cross-section taken along line $n = 200$ of the depth profile of a plane mounted on the translation stage and positioned 10,000 steps from the origin (i.e. $Z = -25.4\text{mm}$). The figure compares graphs calculated from the same acquired images ($s = 32$) using the fixed-point pipeline image processor implementation of the reversed exponential algorithm and a floating-point Matlab implementation. The mean depths calculated by the pipeline processor and Matlab along the cross-section are $-25.4518\text{ mm}$ and $-25.4416\text{ mm}$, respectively. An offset of 0.1 mm has been added to the Matlab result to aid presentation and reduce overlap.
An instrument for real-time profilometry

Figure 4-10: Difference between depth profiles shown in figure 4-9. The mean error introduced by the fixed-point arithmetic implementation is -10.2µm with standard deviation 17.3µm.

Figure 4-11: Photograph of die-cast aluminium pump housing used to demonstrate the real-time system performance.
Figure 4-12: Wrapped phase map showing the machined face of a die-cast aluminium pump component (field of view=120 mm) measured at the maximum fringe density ($s = 32$). The image axes have been scaled to millimetres to aid presentation.

Figure 4-13: Masked depth profile (expressed in millimetres) of the component shown in figure 4-12 calculated using the reversed exponential algorithm implemented on the pipeline image processor. Dark blue regions denote masked pixels. No spatial smoothing or post-image-processing has been performed on the profile.
Figure 4-14: 3-D mesh representation of profile data set shown in figure 4-13. The mesh has been sub-sampled with only one sample in five shown along each axis. Masked pixels are omitted from the mesh.
5 Optimisation of a profilometer based on a spatial light modulator

5.1 Introduction
A digital high-speed high-precision surface profile measuring system implemented using a pipeline image processing development platform was described in chapter 4. A measurement accuracy of better than 1 part in 4,000 was demonstrated using a flat test surface, with the real-time implementation enabling a masked pseudo-colour depth map \((m,n,Z)\) to be displayed after only 0.55 s from the start of the measurement process. However, the fixed-point arithmetic operations and short word lengths used in the real-time implementation compromise the achievable precision of the depth result; for example, the error introduced by truncating the 9-bit subtraction result in stage 1 of the TPUA is quantified in section 3.4. This chapter discusses the results of an experimental study to optimise the measurement precision at the cost of extended measurement and processing times. The study aims to improve the general measurement process for surface profile measuring systems based on spatial light modulators (SLMs) by identifying systematic errors introduced by the hardware arrangement, the phase estimation and phase unwrapping algorithms. In 1995 Biancardi et al. [148] reported on a study of systematic errors relating to profilometers based on fringe projection. The report examines the sensitivity of height map \(h\) to errors in the constant terms \(l_0, d_0,\) and \(f_0\) in equation 1-39 (i.e. spatial dimensions of the experimental layout, and synthetic frequency of the projected fringe pattern); significantly, the report does not discuss errors relating to the estimated phase map \(\Delta \phi\). The study described in this chapter, on the other hand, attempts to identify systematic error sources that affect the measurement and computation of the phase gradient estimator \(\hat{\partial}h\). Investigated issues include: (a) the relative performance of different phase-shifting algorithms (4-frame, 7-frame, and 15-frame); (b) the relative performance of different temporal phase unwrapping algorithms (in particular, the reversed exponential and Fourier transform ranging methods); and (c) the effect of projector defocus.
5.2 Experimental method

The general approach adopted for the study involves projecting a linear sequence of fringe patterns onto the test surface, starting with \( t = 1 \) fringe and increasing linearly in unit increments up to the maximum fringe density \( t = s \) fringes across the projection field. For each fringe density, the fringe pattern is temporally phase-stepped \( M \) times according to the standard phase shifting algorithms described in section 2.2. A SLM based on LCD or DMD technology is able to produce repeatable linear phase steps; phase shifting algorithms designed to reduce estimated phase errors due to non-linearity of the phase-shifting device (e.g. the \( M = N + 1 \) class of algorithms) do not provide any significant performance advantage for systems based on these devices. The study therefore concentrates on the \( M = N \) class of phase shifting algorithms. The standard \( M = N = 4 \) phase shift algorithm, whilst providing a numerically simple means of calculating the unwrapped phase estimate, is sensitive to harmonics at three times the fundamental frequency that are aliased back onto the fundamental. In addition to the four-frame algorithm, therefore, the study evaluates the effect of increasing the number of phase stepped intensity samples per cycle of the fundamental \((N)\), which has a corresponding improvement in the immunity to aliased harmonics. Results are presented for phase shifting algorithms 2, 5, and 6 in table 2-1 (i.e. \( M = N = 4, 7, \) and 15, respectively).

A maximum fringe density of \( s = 64 \) fringes across the projection field was used for the study. As a single EEV CAM17 camera frame occupies approximately 0.25 Mb, the 15-frame phase shifting algorithm results in \( 64 \times 15 \times 0.25 = 240 \) Mb of acquired image data. This clearly exceeds practical storage in random access memory (RAM) and would require significant computational effort to post-process. The study therefore saves only a 128 \times 128 pixel region of interest from the centre of the acquired image to a hard disk drive for post-processing by the mathematical application Matlab, reducing the data set to 15 Mb.

The phase-stepped intensity data is processed offline to produce a sequence of wrapped phase estimate maps. The phase gradient estimator, \( \hat{\phi} \), is computed independently for four of the temporal phase unwrapping algorithms described in section 4.2: (a) unwrapping of a linear sequence of phase values as described by
Optimisation of a profilometer based on a spatial light modulator

equation 4-4, denoted by $Lin$; (b) unwrapping of a linear sequence of phase values and using the intermediate data to fit a straight-line in the least-squares sense as described by equation 4-5, denoted $LinFit$; (c) unwrapping of a reversed exponential sequence of phase values and using the intermediate data to fit a straight-line in a least-squares sense as described by equation 4-11, denoted $RevExp$; and (d) unwrapping using the Fourier ranging method described by equation 4-16, denoted $FT$. The resulting phase gradient estimators are all computed from the same sequence of acquired images and can therefore be quantitatively compared.

Experiments were carried out using the support frame described in section 4.3.3, with the camera, projector, and translation stage in the fixed relative orientation shown in figure 5-1. The distances $L_1$ and $L_2$ (i.e. the distances from the front of the projector and camera to the measurement volume origin) were 485 mm and 725 mm, respectively; $L_3$ was 425 mm, and the projection angle $\alpha = 29^\circ$. The translation stage was used to provide an approximate calibration of the measurement volume, as described in section 4.4. Although accurate to only about 1 part in 1,000, the calibration was nevertheless useful for converting errors in the phase gradient, $\delta\omega$, to equivalent height errors.

The pipeline image processing platform provides a convenient data collection host for the measurement experiments. The digital EEV camera and Proxima projector are both interfaced to the processor and can be synchronised at 60 frames s$^{-1}$. However, it is necessary to identify error sources that are peculiar to this equipment, and where possible minimise these errors, to gain quantitative data describing the precision and robustness of data processing algorithms.

5.3 Projector light source stability issues and SLM non-linearity

Two significant issues arose in connection with the stability of the projector light source. First, a high-frequency ripple in the light output was observed and resulted in fluctuations in the CCD intensity that had a typical standard deviation of 1.2% of the mean intensity at any given pixel (i.e. $-38$ dB). A uniform white scene was projected onto a screen that was imaged by the camera: figure 5-2 shows the mean pixel intensity calculated over one complete camera frame for a sequence of 75 consecutive
frames. A low-frequency envelope modulating the high-frequency error function seems to be present. The discrete Fourier transform (DFT) applied to this data reveals significant frequency components at 18 Hz and 21 Hz (figure 5-3). The proximity of these peaks explains the low frequency envelope function present in the intensity data.

The systematic error function is peculiar to the experimental apparatus and has the potential to mask other error sources that are of more general interest. Since the errors appear to be predominantly sinusoidal, integrating over one complete cycle of the principle observed frequency should cancel the noise superimposed on the pixel intensity. For example, summing $60/[(18 + 21)/2]$ ≈ 3 consecutive frames provides a simple method of improving the signal to noise in the acquired images. A numerical simulation in which the acquired image data was temporally convolved with a rectangular window of variable length revealed that summing six consecutive frames provides the optimum compromise between improved signal to noise ratio and increased acquisition time. Figure 5-4 shows the spatial mean intensity derived from the original signal using this approach and demonstrates a 15 dB improvement in signal to noise ratio.

The second issue involved the fact that instantaneous random changes in the projected light intensity of up to 20% were found to occur at time intervals of the order of minutes. Since the duration of a measurement experiment using the real-time system is less than 1 s this has not been a significant problem. However, the proposed study involves projecting long sequences of fringes, saving the acquired images to disk between consecutive patterns, and will therefore require significantly longer to complete the measurement process. For example, given that it takes approximately 0.5 s to save the acquired data for one fringe pattern to the hard drive, an experiment using a linear sequence of fringes up to a fringe density $s = 128$ fringes and a $N = M = 4$ phase stepping scheme will take approximately 256 s.

Extending the duration of the measurement process in this way increases the probability that a random change in intensity will occur during the experiment requiring that the experiment be repeated. Furthermore, repeating experiments of longer duration soon becomes very time consuming.
Optimisation of a profilometer based on a spatial light modulator

The random intensity changes were attributed to ageing of the bulb since the problem disappeared on replacing it with a new one, only to reappear after around 100 h of use (only 1/10th of the specified MTTF lifetime of the bulb). With a replacement OSRAM metal-halide 270 W bulb costing around £400 this problem may be an issue for a practical shape measurement system based on projectors of this type. The solution for these experiments was to use the pipeline processor hardware to calculate the mean intensity of the acquired images in real-time, and to re-acquire the set of phase shifted images at a given \( t \) value if this changed by more than a certain threshold value for one of the frames.

The non-linear variation of the projected light intensity was characterised by projecting a sequence of uniform images with increasing intensity and plotting the spatial mean intensity in the images acquired using the CCD camera (figure 5-5). The experiment was repeated with a range of neutral density filters (\( \eta = 2, 4, 8 \)) inserted in front of the camera lens to reduce the dynamic range of the signal applied to the CCD array and the transfer function was found to scale approximately linearly against \( 1/\eta \): the scaling is not exactly linear due to the random changes in projector bulb intensity between experiments. Since the transfer function appears to be independent of the CCD dynamic range, the non-linearity can be attributed to the projector SLM rather than the CCD, and closely follows the non-linearity of figure 4-4. The inverse transfer function can be computed and used to modify the projected fringe patterns, as carried out in the real-time implementation (section 4.5.5). Significantly, we conclude that this correction can be applied in all experiments independently of the camera settings (lens, aperture, neutral density filter, etc.).

5.4 Effect of SLM pixel structure

This section presents experimental results that illustrate the effect of the SLM pixel structure on the measurement depth profiles. The possibility of reducing the resulting errors by deliberately defocusing the projector is investigated, and an optimal defocus distance is found to exist at which the errors are minimised.

The experiments were carried out on a flat glass plate spray-coated with matt white emulsion paint. The projector was driven at an effective resolution of 400 horizontal x
600 vertical pixels. A horizontal cross section of the profile, calculated using unwrapping method LinFit with \( s = 32 \), is shown in figure 5-6. In this experiment, the projector was focused on the plate, which was positioned at the centre of the measurement volume. The indices \( m \) and \( n \) are used to denote horizontal and vertical camera pixel co-ordinates, respectively. The \( m \) dependence of the profile was removed by least-squares fitting a fourth-order polynomial in \( m \); the best-fit curve \( \hat{\omega}_f(m) \) was then subtracted from \( \hat{\omega}(m) \). A 128 pixel line of the resulting error values \( \Delta \hat{\omega}(m) \), obtained from three plots similar to the central region of the cross section in figure 5-6, is plotted in figure 5-7 for unwrapping methods Lin, LinFit, RevExp, and FT. In the centre of the measurement volume, an error \( \Delta \hat{\omega} \) in the phase gradient equates to an error in the measured depth \( \Delta Z \) given by

\[
\Delta Z = A_1 \Delta \hat{\omega},
\]

where constant \( A_1 \) was determined from the calibration experiment as \( A_1 = 0.0753 \text{ m rad}^{-1} \). Substituting a value \( \Delta \hat{\omega} = 4.61 \times 10^{-4} \) (equal to the standard deviation of the fluctuations for method FT) gives a depth uncertainty \( \Delta Z = 35 \text{ µm} \).

Throughout this section we use the value of \( \Delta Z \) calculated by equation 5-23 from the standard deviation of the \( \hat{\omega}(m) - \hat{\omega}_f(m) \) cross sections to indicate the height precision that has been achieved. If the system had been fully calibrated, then this precision value would have corresponded to a one-sigma height measurement error.

The profile of the plate was measured independently by Dr. Philip Aldred at Brown and Sharpe Ltd using a point-wise laser scanner. The plate was scanned along several sections and the main feature was found to be a waviness of only 2 µm peak-to-peak amplitude with a period of the order of 1 mm. The high-frequency oscillations giving rise to the majority of the error in the profile result obtained using the fringe projection system cannot therefore be due to intrinsic features of the test surface. Furthermore, these errors are not due to some artefact of the camera since changes in the camera magnification result in changes to the spatial frequency of the oscillations. The factor responsible is the finite size of the projector pixels. When the projector is focused on the plate, the calculated phase gradient will vary in a step-like fashion when crossing from one SLM pixel image to the next, instead of the continuous linear variation expected for pure sinusoidal fringes. It is clear from this interpretation that
increasing the maximum number of fringes $s$ will not improve matters, since the quantisation of the $\dot{\omega}(m)$ profile will remain.

The solution to the problem adopted here was to deliberately defocus the projector so as to low-pass filter the projected fringes. If the blurring is too severe, however, the modulation of the higher frequency fringes will also be attenuated, leading to increased depth measurement errors. The remainder of this section describes the results of experiments to find the location of the optimal focus plane position at which the overall depth measurement error is minimised.

The position of the projector focal plane was adjusted to one of nine sample distances behind the surface of the glass flat. The projector focal plane distance was determined by projecting a high-frequency vertical grid into the measurement volume, which enabled the focus distance to be determined to an accuracy of approximately 20 mm to 30 mm. At each focal distance a linear sequence of fringe patterns ($1 \leq t \leq 64$) was projected for each of the three phase shifting methods: $M = N = 4, 7, \text{ and } 15$; and the acquired data was analysed using each of the five temporal phase unwrapping methods.

Figures 5-8(a) to 5-8(c) illustrate the variation of the standard deviation of $\dot{\omega}(m) - \dot{\omega}_r(m)$ with the distance between test surface and the plane of best focus, for three different phase shifting algorithms ($M = N = 4, 7, \text{ and } 15$) and phase unwrapping method $FT$. The values were calculated from the mean of 32 independent lines in the image (each separated from the next by 3 lines), and error bars show the standard deviation of the mean. An optimal range for the projector focal plane distance is between 230 mm and 460 mm behind the test surface. Within this range, the observed errors are reduced by increasing the number of projected fringes from $s = 16$ to $s = 32$. However, no significant improvement is achieved by further increments to $s = 64$. Comparison of the three figures shows that there is also no significant improvement in the observed phase gradient error when the number of phase steps is increased from 4 to 15. Substituting $\Delta \dot{\omega} = 1.3 \times 10^{-4}$ into equation 5-23 shows that the optimum measurement precision achieved using the Fourier ranging method is approximately 9.8 $\mu$m ($1/15380$ of the measurement volume side length).
This value corresponds to the case $M = N = 4$ and $s = 64$, with the projector focal plane position 230 mm behind the test surface.

Figures 5-9(a) to 5-9(c) show the results of post-processing the data sets using phase unwrapping method $RevExp$. The reversed exponential method is not quite able to match the precision achieved using the Fourier ranging method, but nevertheless the loss in performance is relatively small and probably acceptable in many situations where the short measurement and analysis times make it an attractive approach. Substituting $\Delta \phi = 1.5 \times 10^{-4}$ in equation 5-23 indicates an optimal measurement precision of approximately 11 µm (1/13300 of the measurement volume side length). This was achieved with $M = N = 4$ and $s = 32$, with the projector focal plane positioned 230 mm behind the test surface.

Experiments were subsequently carried out with the SLM operating at its full resolution of 800 horizontal $\times$ 600 vertical pixels. Doubling the number of horizontal pixels was found to reduce the depth error to around 7.4 µm (Fourier ranging method $FT; N = M = 4; s = 64$; focal distance 150 mm behind the test surface), which corresponds to a measurement precision of approximately 1 part in 20,000 of the measurement volume side length. The results obtained using the reversed exponential unwrapping method also provided a comparable precision figure.

5.5 Repeatability

Repeatability is a key issue for a commercial shape measurement system. Experiments were therefore carried out to assess the changes in measured depth values for a nominally stationary component occurring over time scales of up to four hours. The high temperatures and presence of three separate cooling fans in the projector makes this the most likely source of drift over time. Motion of the SLM perpendicular to the optic axis of the projector results in a uniform phase change of the projected fringes, causing an apparent rigid body motion of the sample. Motion of the SLM along the optic axis causes a magnification change, which is probably more serious in many applications since the measured dimension of a component also changes by the same factor.
The test object for this experiment was a stepped aluminium block mounted on a second aluminium block to give a total height range of 108.9 mm. The surface profile measured at the start of the experiment, $T = 0$, is represented in figure 5-10. The camera pixel column and row values ($m$ and $n$, respectively) have been used as spatial co-ordinates in the 3-D plot, rather than real-world $X$ and $Y$ co-ordinates, to improve the clarity of the following discussion. Over the 4 h time period, the top and bottom steps were found to undergo an apparent absolute displacement along the $Z$-axis of up to 130 µm and 140 µm, respectively. Figure 5-11 illustrates the repeatability of the absolute $Z$ values measured across the bottom step along a cross section through row $n = 10$. To assess the performance of relative $Z$ measurements, the height of each step was calculated with respect to the bottom step. One such cross section from each experiment is plotted in figure 5-12. Small variations in the relative $Z$ values were found to occur with increasing distance from the bottom step, although these are not visible in figure 5-12 due to the wide height range of the vertical axis. Cross sections across the top step relative to the bottom step, measured in a perpendicular direction to those for figure 5-12 (i.e. parallel to the step edges), are shown in figure 5-13 on an expanded vertical scale. The calculated relative $Z$ values vary over a range of 180 µm, which represents a significant drift with respect to the measurement baseline of 108.9 mm (i.e. 1 part in 605). Analysis of the relative height of each step on the stepped block revealed that the range of drift scaled linearly with the $Z$ value. The calculated magnification factor $M(T)$ is shown in figure 5-14.

Correction of the relative $Z$ data in each experiment was carried out by dividing by $M(T)$; the curves in figure 5-13 were then found almost to collapse onto a single line, as shown in figure 5-15. The residual time variation of the corrected relative $Z$ values at a given pixel can be quantified by the standard deviation $\sigma_Z$, which is plotted in figure 5-16 for the data from figure 5-15. The average value of $\sigma_Z$ is 6.5 µm, which is comparable to the precision values calculated in section 5.4. It is clear from figure 5-15 that the small-scale structure of the measured profile is highly reproducible from measurement to measurement, indicating that random errors are not the cause of the spatial fluctuations. Spectral analysis of the traces shown in figure 5-15 resulted in a significant peak at the spatial frequency of the SLM pixels, indicating that the residual
Optimisation of a profilometer based on a spatial light modulator

Pixel structure effects are still the dominant measurement error source, even though the experiment was carried out at the optimum focal plane setting of the projector.

5.6 Over-exposure of image sensor

When profiling optically smooth surfaces, the variation between the maximum and minimum reflected intensity values can be many orders of magnitude. The limited dynamic range of most image sensors (typical 8-bit devices have a maximum dynamic range of 255:1) means that both under- and over-exposure are likely to occur quite frequently in practice. Under-exposure decreases the fringe modulation and hence increases the random error in the measured depth values[144], whereas saturation is a form of non-linear detector response, which is well known to introduce systematic errors into the estimated phase. The purpose of this section is to present experimental results demonstrating the susceptibility of various phase shifting and phase-unwrapping algorithms to these errors.

The glass flat used in section 5.4 provided a uniform surface finish. The camera aperture was initially set at f/15, which provided the maximum image intensity without exceeding the dynamic range of the pixel sensors. The experiment was carried out with the camera aperture diameter increased by approximately one stop (f/11) to over-expose some of the CCD pixels. Phase gradient errors produced by the four-frame and seven-frame phase-shifting algorithms and all four phase unwrapping algorithms were then compared.

Figure 5-17 shows a horizontal cross-section of the measured intensity for one of the images acquired with \( t = 32 \) projected fringes and f/11. The non-uniform profile is a consequence of the fact that the glass flat is mounted perpendicular to the optic axis of the camera and illuminated obliquely by the projector. Saturation is clearly occurring for pixel co-ordinate \( m \) greater than about 300.

Phase gradient error plots are shown in figure 5-18 for data analysed by the four-frame phase-shifting method and using the four phase unwrapping methods. Increased errors are clearly visible in the over-exposed region. The basic phase unwrapping method (Lin) is the most sensitive to the clipped intensity values, resulting in errors up to an order of magnitude greater than those from the Fourier ranging method (FT)

150
calculated using the same data. Although not as sensitive as the Lin method, the reversed exponential method (RevExp) introduces significant errors when presented with clipped intensity data.

Measurements obtained using the seven-frame phase shifting method with the same camera aperture are illustrated in figure 5-19. Significant reductions in the phase gradient error (compared with the four-frame method) were found with the basic temporal unwrapping algorithm (Lin). This is to be expected since the temporal sampling rate of the four-frame method causes the harmonic at three times the fundamental frequency to be aliased directly onto the fundamental, whereas for the seven-frame algorithm, the first harmonic to cause phase-extraction errors is that at six times the fundamental frequency[18]. The reversed exponential temporal phase unwrapping algorithm and the Fourier ranging method also show reduced errors compared to the same methods with the four-frame data, although the improvement is not so dramatic as with the basic temporal phase unwrapping method (Lin).

The aperture was opened to f/8 causing significant clipping of the acquired images as shown in figure 5-20. The phase gradient error plots corresponding to the four- and seven-frame phase shifting algorithms are shown in figures 5-21 and 5-22, respectively. The four-frame phase shifting algorithm results in significant errors in the phase gradient estimate along most of the cross-section using the linear (Lin) and reversed exponential (RevExp) methods. The error function is modulated by a low-frequency envelope, which is minimised when the acquired intensity signal is clipped at 50% of its peak-to-peak amplitude. The phase gradient errors observed using the linear method are approximately one order of magnitude greater than those calculated with an optimised camera aperture using the same unwrapping method. The Fourier ranging (FT) and linear least-squares fitting (LinFit) unwrapping methods continue to demonstrate better immunity to the clipped intensity data, however, significant errors are now becoming visible. It is interesting to note that the phase gradient error function for the latter two unwrapping methods agree very closely. Figure 5-23 shows the spatial frequency spectrum for the clipped intensity data with harmonics of the fundamental fringe frequency clearly present. A similar analysis of the $N(t)$ signal (figure 5-24) and $D(t)$ signal calculated using equations 2-27 and 2-28 reveal that
Optimisation of a profilometer based on a spatial light modulator

these higher harmonics are transmitted intact by the four-frame phase shifting algorithm, although somewhat attenuated. Phase extraction using the seven-frame phase shifting algorithm results in reduced errors in the phase gradient estimate compared with the four-frame method. Applying the discrete Fourier transform (DFT) to the $N(t)$ signal (figure 5-25) and $D(t)$ signal computed using the seven-frame phase shifting algorithm shows that these higher harmonics have been significantly attenuated.

It is clear, however, that in general over-exposure is a serious error source that must be addressed when implementing a practical measurement system. Detection of saturating pixels could be achieved by simply masking those pixels whose measured intensity values reach 255 grey levels in any one of the $M$ acquired frames, and indeed this method is implemented in the high-speed profilometry system described in chapter 4. The effective dynamic range of the projector-camera system could also be increased by controlled modulation of the SLM transmittance, such that projected intensity is reduced in regions of high reflectivity[147]. Depending on the end application of the system, it may be worth considering a CCD camera with an effective dynamic range of 12 bits (4096:1) or higher.

5.7 Conclusions

The performance of a shape measurement system, based on projected fringes generated by an SLM, was assessed using a systematic set of experiments. Three phase-shifting algorithms, five temporal unwrapping algorithms, nine projector defocus distances, and two SLM resolutions were studied. These have confirmed that the measurement precision depends critically on the position of the projector’s focal plane relative to the measurement volume. The attainable precision represents a compromise between the systematic errors introduced by the pixel structure of the projector, and the random errors introduced by electronic noise. The defocus required to reduce the systematic errors sufficiently attenuates the fringe modulation at high fringe densities to such an extent that the theoretical advantages offered by large $s$ values cannot be realised in practice. Nevertheless, even using $s = 64$, a measurement precision (at the 1 $\sigma$ level) of 1 part in 20,000 of the measurement volume side length was achieved.
Optimisation of a profilometer based on a spatial light modulator

Repeatability experiments carried out over a time scale of 4 h showed significant changes in both the absolute and relative measured co-ordinates, which are consistent with a fractional change in projector magnification within ±0.1% over this time period. When the magnification factor was calculated from the data, all relative cross-sections collapsed onto a single line, indicating that provided care is taken over the optical system design and implementation to ensure sufficient stability, repeatability values of less than 10 µm should be attainable.

Over-exposure of the CCD array by a relatively small factor (2 times) causes large increases in measurement error (factors of up to an order of magnitude). This can be reduced by choosing a phase-shifting algorithm that is less sensitive to harmonics caused by clipping of the acquired intensity data.
5.8 Figures

Figure 5-1: Layout of the experimental apparatus used to acquire the data in this chapter.

Figure 5-2: Spatial mean intensity calculated for a sequence of 75 consecutive EEV CAM17 camera frames; a uniform white scene was projected by the Proxima DP4200 on to a screen that was imaged by the camera. Temporal mean intensity is 165.2 with standard deviation 2.0 grey levels.
Figure 5-3: One-dimensional discrete Fourier transform (DFT) applied to the spatial mean intensity signal shown in figure 5-2. Mean intensity was subtracted before DFT to remove the DC peak.

Figure 5-4: Convolution of spatial mean intensity signal shown in figure 5-2 with a rectangular function to simulate temporal integration over six frame periods. The standard deviation is 0.3 grey levels and has been reduced by 15 dB compared to the original signal.
Optimisation of a profilometer based on a spatial light modulator

Figure 5-5: Spatial mean pixel intensity in acquired image plotted against the projector pixel intensity. Graphs were obtained both with, and without, a neutral density filter placed in front of the camera lens. Results are presented for neutral density factors 2, 4 and 8, corresponding to $\eta=2$, 4, and 8, respectively.

Figure 5-6: Cross section of measured phase gradient distribution for a flat plate computed using the LinFit phase unwrapping method and four-frame phase shifting algorithm ($M=N=4$). The maximum fringe density was $s=32$. 
Figure 5-7: Central portion of phase gradient error curve computed using four-frame phase shifting algorithm (M=N=4) and maximum fringe density s=16. The plane of best focus is coincident with the test surface. Graphs for LinFit, RevExp, and FT phase unwrapping methods are offset by $2 \times 10^{-3}$, $4 \times 10^{-3}$, and $6 \times 10^{-3}$ rad, respectively.
Figure 5-8: Mean value of the standard deviation of the phase gradient error $\sigma_\omega$ versus projector defocus distance, calculated using Fourier ranging method (FT). Data for $s=16$, $s=32$, and $s=64$ are plotted in the line styles indicated; (a), (b), and (c) show the results for phase shifting algorithms $M=N=4$, $7$, and $15$, respectively.
Figure 5-9: Mean value of the standard deviation of the phase gradient error $\sigma_{\omega}$ versus projector defocus distance, calculated using reversed exponential unwrapping method (RevExp). Data for $s=16$, 32, and 64 are plotted in the line styles indicated; (a), (b), and (c) show the results for $M=N=4$, 7, and 15, respectively.
Figure 5-10: Profile of the stepped test surface measured at time $T=0$. No smoothing has been applied to the data, however, the three-dimensional mesh shows only one sample point in four along each axis (i.e. $1/16^{th}$ of the complete data set). Pixels at which the measured normalised fringe modulation fell below 0.1 are omitted from the result. The axes are scaled according to the real-world dimensions.

Figure 5-11: Repeatability of absolute $Z$ calculated along row $n=10$; the cross section runs through the bottom step in figure 5-10 (i.e. step furthest from the camera) parallel to the $X$-axis. The absolute measurement varies over a range of 140µm during a period of 240 minutes.
Optimisation of a profilometer based on a spatial light modulator

Figure 5-12: Repeatability of relative Z calculated with respect to the bottom step along column $m=10$; the cross section runs parallel with the Y-axis.

Figure 5-13: Repeatability of relative Z calculated along $n=340$; the cross section runs through the top step in figure 5-10 (i.e. step nearest to the camera) parallel to the X-axis. The relative profile varies over a range of 180µm during a period of 240 minutes.
Optimisation of a profilometer based on a spatial light modulator

Figure 5-14: Magnification correction factor $M(T)$ for relative $Z$ calculated by comparing distance between top and bottom step for each experiment.

Figure 5-15: Repeatability of relative $Z$ corrected using coefficient $M(T)$ calculated along $n=340$; the cross section runs through the top step in figure 5-10 parallel to the $X$-axis. The scale of the abscissa axis is chosen to match figure 5-13 for comparison.
Figure 5-16: Standard deviation of the corrected Z values displayed in figure 5-15, calculated over time for each pixel.

Figure 5-17: Effect of sensor saturation: horizontal cross section through acquired image using camera aperture f/11 (r=32 fringes).
Figure 5-18: Measured phase gradient error using camera aperture f/11 for each of the phase unwrapping methods. Results obtained using maximum fringe density $s=32$, with the four-frame phase shifting algorithm ($M=N=4$). The LinFit, RevExp and FT graphs have been offset by 0.004, 0.008, and 0.012 rad, respectively.

Figure 5-19: Measured phase gradient error. Details are identical to figure 5-18 except that the data was obtained using the seven-frame algorithm ($M=N=7$).
Optimisation of a profilometer based on a spatial light modulator

Figure 5-20: Effect of sensor saturation: horizontal cross section through acquired image using camera aperture f/8 (r=32 fringes).

Figure 5-21: Measured phase gradient error using camera aperture f/8 with the four-frame phase shifting algorithm \((M=N=4)\). Results obtained using maximum fringe density \(s=32\). The LinFit, RevExp and FT graphs have been offset by 0.004, 0.008, and 0.012 rad, respectively. The scale of the abscissa axis is chosen to match figure 5-18 for comparison.
Optimisation of a profilometer based on a spatial light modulator

Figure 5-22: Measured phase gradient error. Details are identical to figure 5-21 except that the data was obtained using the seven-frame algorithm ($M=N=7$).

Figure 5-23: One-dimensional discrete Fourier transform (DFT) of the acquired image intensity data shown in figure 5-20. The mean intensity was subtracted prior to applying the DFT to remove the DC-peak and the data was zero padded by a factor of 20 to improve the resolution along the spatial frequency axis.
Figure 5-24: One-dimensional discrete Fourier transform (DFT) of the $N(t)$ signal, which is proportional to the sine of the wrapped phase. The $N(t)$ signal is calculated according to equation 2-27 and using the four-frame phase shifting algorithm ($M=N=4$) sampling coefficients. The mean signal value was subtracted prior to applying the DFT to remove the DC-peak and the data was zero-padded by a factor of 20 to improve the resolution along the spatial frequency axis.
Figure 5-25: One-dimensional discrete Fourier transform (DFT) of the \( N(t) \) signal calculated using the seven-frame phase shifting algorithm \( (M=N=7) \) sampling coefficients. The mean signal value was subtracted prior to applying the DFT to remove the DC-peak and the data was zero-padded by a factor of 20 to improve the resolution along the spatial frequency axis.
Section IV:
Temporal phase unwrapping algorithm applied to digital speckle pattern interferometry
6 An instrument for real-time digital speckle pattern interferometry

6.1 Introduction

Whole-field digital speckle pattern interferometry techniques introduced in chapter 1 enable measurement of the deformation of rough surfaces to be made with high sensitivity. The procedure used to analyse the resulting phase-stepped two-dimensional interferograms (explained in detail in chapter 2) has much in common with the analysis of interferograms obtained using the real-time profilometry system presented in chapter 4. This chapter describes the design of a quantitative analysis system for digital speckle pattern interferometry based on pipeline image processing that enables two-dimensional unwrapped phase maps to be displayed in real-time. The system complements and extends the real-time profilometry system, allowing the user to perform profilometry and DSPI measurements in direct spatial registration from within the same software application.

Interferograms formed by the interference of two speckle wavefronts, or one speckle wavefront and a smooth reference wavefront, introduce some additional problems that are not significant in the case of white light projected fringe profilometry, and practical steps are required to reduce their effect on the phase estimate. Speckle fields produced by the illumination of rough surfaces are characterised by randomly varying intensity and phase values. The random phase of each speckle prevents the formation of fringes in the classical sense, and it is therefore necessary to subtract the random phase corresponding to initial and final states to yield a phase map representing the physical quantity of interest. In phase-shifting speckle interferometry, the varying intensity of speckle fields always results in the presence of both poorly-modulated and well-modulated pixels in the speckle interferogram. Unfortunately, saturation or low modulation of the photodetector prevents reliable extraction of the phase estimate by phase-shifting techniques and so care must be taken in adjusting the wavefront intensities. In phase-shifting speckle interferometry each point in the image is considered as an individual interferogram. If the speckles are sufficiently small to be
An instrument for real-time digital speckle pattern interferometry

perfectly resolved by the CCD elements then each pixel behaves as a classical interferometer with its individual background intensity and modulation. A large number of the discussions appearing in the literature offering theoretical optimisations of speckle interferometer arrangements are based on this assumption, and require that the lens aperture be reduced such the speckle size approaches the pixel dimension. For example, based on statistical analysis of camera resolved speckles, Lehmann[149] suggests that the average reference wavefront intensity should not exceed 25% of the detector saturation level for smooth-reference speckle interferometers, with the average speckle wavefront intensity adjusted to the range 5 to 10%. Huntley[150] notes that the speckle wavefront intensity has little effect on the observed phase error distribution, and derives models to describe the error distribution as a function of the number of quantisation levels in the photodetector, and the fraction of pixels that are allowed to saturate. In the case of two superimposed speckle fields, Lehmann suggests that the average speckle wavefront intensities should each be adjusted to approximately 15% of the saturation level. However, it is often preferable to work with a larger lens aperture that allows the power of the illuminating beam to be reduced and/or the use of shorter camera exposure times. In this case, the speckle size is reduced such that several speckles now fall within the area of an individual pixel. Lehmann[99] more recently noted that speckle interferometry using unresolved speckles can in fact produce good results provided care is taken in adjusting the wavefront intensities with respect to the photodetector saturation level as described above and that the minimum modulation required by the phase estimate calculation is not too high. The resulting background intensity measured at each pixel corresponds to the average of the different background intensities of the speckles within the pixel:

$$I_0 = \frac{1}{n} \sum_{k=1}^{n} I_{0k},$$

(6-1)

where \(n\) is the number of speckles falling within a given pixel. The observed modulation and phase satisfy

$$I_m \exp(i\phi) = \frac{1}{n} \sum_{k=1}^{n} I_{mk} \exp(i\phi_k),$$

(6-2)

where the uniform distribution of the phases \(\phi_k\) therefore leads to a lower measured modulation than that of the individual speckles. A further paper by Lehmann[151] examines the effect of changes in the position and aspect of speckles during the
An instrument for real-time digital speckle pattern interferometry

measurement period, so-called *speckle decorrelation*, and determines that
decorrelation-induced phase errors are independent of the number of speckles per
pixel and therefore behave exactly the same for both resolved and unresolved speckle
interferograms. Furthermore, the result is the same regardless of whether the reference
field is a smooth wavefront or a speckle wavefront. The decorrelation-induced phase
errors are determined only by the amount of decorrelation, $\rho$, and the pixel
modulation, $I_M$. The amount of decorrelation is given by the ratio of the pixel
dimension (or lens diameter in the case of pupil plane decorrelation) replaced during
decorrelation, and is reduced for large pixels in the case of image plane decorrelation
and for large lens apertures in the case of pupil plane decorrelation. In each case, the
number of speckles per pixel will tend to increase.

6.2 Extending the temporal phase unwrapping algorithm for DSPI

6.2.1 Combining independent phase estimates

In addition to the practical steps outlined in the preceding section for reducing phase
errors due to speckle phenomena in phase-stepping interferometers, the basic
temporal phase unwrapping algorithm should be extended to improve the phase
estimate. Errors in the phase estimate due to electronic noise and small amounts of
speckle decorrelation can be approximated by a Gaussian distribution with variance
inversely proportional to the square of the pixel modulation[151]. Assuming that the
phase-change function is slowly-varying, the signal to noise ratio can be improved by
combining independent temporally adjacent phase-change samples measured at the
same pixel, or spatially adjacent samples measured at the same time. An example of
the former method is the so-called temporal speckle averaging technique in which a
sequence of images proportional to $\sin \Delta \hat{\phi}_w$ and $\cos \Delta \hat{\phi}_w$ are obtained using different
object beam illumination angles, and then summed[152]. Section 2.4.2 outlined an
efficient arithmetic implementation of the latter method, so-called spatial speckle
averaging, that combines spatially adjacent phase estimates by convolution of the
$F(t_1,t_2)$ and $G(t_1,t_2)$ images (defined by equations 2-50 and 2-51, respectively),
proportional to $\sin \Delta \hat{\phi}_w$ and $\cos \Delta \hat{\phi}_w$, respectively, with a square filter function. For
many applications spatial speckle smoothing is the more convenient approach for
An instrument for real-time digital speckle pattern interferometry

achieving improved signal to noise ratio, at the expense of reduced spatial resolution. The problem remains, however, to select the optimum sample weightings that produce the best estimate of the phase-change based on the combined samples.

A common approach to such problems involves modelling the physical system with a set of parameters $P$, and then given a sample $S$ of observed data, selecting the value that maximises the likelihood function $L(P | S)$ as the best estimate. In the case of a phase-stepped speckle interferometer the sample values are the complex amplitudes $a_o(t_1)$ and $a_o(t_2)$ observed at times $t_1$ and $t_2$, respectively. Figure 6-1 shows a phasor diagram representation of complex amplitudes measured at a single pixel in the image plane before and after a change in the physical parameter of interest. The observed final complex amplitude $a_o(t_2)$ is regarded as consisting of the initial observed amplitude, $a_o(t_1)$ rotated by $\Delta \phi$ due to a change in the physical parameter, plus a noise term $a_n$ that represents the error introduced by electronic noise and speckle decorrelation:

$$a_o(t_2) = Ra_o(t_1) + a_n, \quad (6-3)$$

where

$$R = \begin{bmatrix} \cos \Delta \phi & -\sin \Delta \phi \\ \sin \Delta \phi & \cos \Delta \phi \end{bmatrix} \quad (6-4)$$

The phase change value of interest $\Delta \phi$ therefore differs from the observed value $\Delta \phi$ due to the noise term. The vectors can be expressed in Cartesian form (dropping the time index for simplicity),

$$a_o = \begin{bmatrix} a_o^x \\ a_o^y \end{bmatrix}, \quad (6-5)$$

and if we assume the usual first order speckle statistics[153] then $a_o^x$, $a_o^y$ are independent Gaussian random variables. A sample of $Q$ data values measured by a phase-stepping interferometer can be expressed as $a_o(t_1)$ and $a_o(t_2)$ where the additional subscript $i$ refers to the $i$th independent measurement of the speckle phase change. The Cartesian components of the corresponding noise vector $a_n$ are independent Gaussian random variables with standard deviation $\sigma_n$ and are given by
An instrument for real-time digital speckle pattern interferometry

\[
\alpha_{\text{ii}} = \alpha_{\alpha\alpha}(t_2) - \alpha_{\alpha\alpha}(t_1) \cos \Delta \phi + \alpha_{\alpha\beta}(t_2) \sin \Delta \phi, \quad (6-6)
\]

\[
\alpha_{\text{ii}} = \alpha_{\alpha\alpha}(t_2) - \alpha_{\alpha\alpha}(t_1) \sin \Delta \phi - \alpha_{\alpha\beta}(t_2) \cos \Delta \phi, \quad (6-7)
\]

The physical system is therefore fully described by the parameters \( \Delta \phi \) and \( \sigma_n \), with the likelihood function given by

\[
L(P | S) \propto \prod_{i=1}^{Q} \frac{1}{2\pi\sigma_n^2} \exp \left\{ -\frac{\left[ \alpha_{\alpha\alpha}(t_2) - \alpha_{\alpha\alpha}(t_1) \cos \Delta \phi + \alpha_{\alpha\beta}(t_2) \sin \Delta \phi \right]^2}{2\sigma_n^2} \right\}
\]

(6-8)

Huntley[150] shows that this leads to the maximum likelihood estimator for \( \Delta \phi \) as

\[
\hat{\Delta \phi} = \tan^{-1} \left[ \frac{\sum_{i=1}^{Q} F(t_1, t_2)}{\sum_{i=1}^{Q} G(t_1, t_2)} \right] \quad (6-9)
\]

The best estimate of the phase-change, in the maximum likelihood sense, for \( Q > 1 \) is therefore obtained by averaging the \( F(t_1, t_2) \) and \( G(t_1, t_2) \) images, with no additional sample weighting, before carrying out the arctangent operation.

An alternative study by Lehmann[151] calculates the optimum phase-change estimator as a weighted spatial average of independent \( \Delta \phi_{\text{w}, i} \) values using weightings that are inversely proportional to the variance of the error distributions, \( \sigma_i \),

\[
\hat{\Delta \phi} = \left[ \sum_{i=1}^{Q} \frac{1}{\sigma_i^2} \right]^{-1} \sum_{i=1}^{Q} \frac{\Delta \phi_{\text{w}, i}}{\sigma_i^2}, \quad (6-10)
\]

where the phase-change values are calculated as,

\[
\Delta \phi_{\text{w}, i} = \tan^{-1} \left[ \frac{\alpha_{\alpha\alpha}(t_2)}{\alpha_{\alpha\alpha}(t_1)} \right] - \tan^{-1} \left[ \frac{\alpha_{\alpha\beta}(t_2)}{\alpha_{\alpha\beta}(t_1)} \right] \quad (6-11)
\]

The variance of the error distributions in this case is found to be inversely proportional to the product of the modulation at each pixel before and after deformation[151]:

\[
\sigma_i^2 \propto \frac{1}{I_{\alpha\alpha}(t_1)I_{\alpha\alpha}(t_2)}, \quad (6-12)
\]

giving the maximum likelihood estimator.
An instrument for real-time digital speckle pattern interferometry

\[
\Delta \phi_w = \left[ \sum_{i=1}^{Q} I_{M_1}(t_i)I_{M_1}(t_2) \right]^{-1} \sum_{i=1}^{Q} I_{M_1}(t_i)I_{M_1}(t_2) \Delta \phi_{w,i} \quad (6-13)
\]

Whilst evaluation of the estimate having maximum likelihood in each of the above methods differs, the weightings used to combine the independent samples are the same in each case. This can be seen, for example, if we consider the standard four-frame phase stepping algorithm \((M=N=4)\), where equation 6-9 can now be expanded by substituting equations 2-7 and 2-52, to give

\[
\Delta \hat{\phi}_{w} = \tan^{-1} \left[ \frac{\sum_{i=1}^{Q} 4I_{M_1}(t_i)I_{M_1}(t_2) \sin \Delta \phi_{w,i}}{\sum_{i=1}^{Q} 4I_{M_1}(t_i)I_{M_1}(t_2) \cos \Delta \phi_{w,i}} \right] \quad (6-14)
\]

It can be seen from equation 6-14 that spatial averaging of the \(F(t_1,t_2)\) and \(G(t_1,t_2)\) images obtained using the four-frame algorithm implicitly applies weightings to the independent phase-change samples that are proportional to the product of the modulation at times \(t_1\) and \(t_2\). In general, this can be demonstrated for all standard phase stepping algorithms. However, there is an important difference between the two methods in that equation 6-13 will give a different result when wrapped \(\Delta \phi_{w,i}\) samples (i.e. in error by \(2\pi\)) are combined, whereas equation 6-14 will not. This has significant practical implications for the signal to noise ratio in the presence of unwrapping errors as is demonstrated in section 6.2.4.

6.2.2 Updating the reference phase

The modified sum of differences method of temporal phase unwrapping implemented in the high-speed profilometry system and described in section 2.4.3 calculates the phase-change values with respect to the phase estimate at time \(t=0\) (i.e. \(\Delta \hat{\phi}(t,0)\)). This has the advantage that intermediate phase values only contribute to the calculation of the integer valued correcting field \(v(t)\) and therefore small errors in these intermediate values are not integrated into the final unwrapped phase. However, in speckle interferometry applications significant deformation of the test surface may result in complete speckle decorrelation between the initial and final states[154]. Speckle decorrelation is primarily observed due to: (a) in-plane motion of the speckle at the objective lens due to specimen tilt and strain; (b) in-plane motion of the speckle
An instrument for real-time digital speckle pattern interferometry at the image plane due to in-plane motion of the specimen; and (c) changes in the scattering properties of the specimen surface. In such cases, the speckle reference pattern from $t = 0$ should be updated periodically, at times $t = r_1, r_2, \ldots, r_k$, such that only a relatively small amount of decorrelation occurs within any one given time increment. The total wrapped phase change over $(\kappa + 1)$ incremental phase maps will then be as follows:

$$\Delta \hat{\phi}(t, 0) = \Delta \hat{\phi}(t, r_k) + \sum_{k=2}^{k} \Delta \hat{\phi}(r_k, r_{k-1}) + \Delta \hat{\phi}(r_1, 0)$$

(6-15)

where the incremental unwrapped phase difference is obtained as,

$$\Delta \hat{\phi}(t_1, t_2) = \Delta \hat{\phi}_w(t_1, t_2) - 2\pi \nu(t_1, t_2)$$

(6-16)

with the total number of phase jumps occurring between time $t_2$ and $t_1$ calculated as

$$\nu(t_1, t_2) = \sum_{t'=t_1}^{t_2} d(t', t_2), \text{ for } t_1 > t_2$$

$$\nu(t_1, t_2) = 0, \text{ for } t_1 = t_2$$

(6-17)

where

$$d(t_1, t_2) = \text{NINT} \left\{ \frac{\Delta \hat{\phi}_w(t_1, t_2) - \Delta \hat{\phi}_w(t_1 - 1, t_2)}{2\pi} \right\}$$

(6-18)

The optimum interval between reference speckle pattern updates will depend on the rate of speckle decorrelation; the special case of re-referencing every processing frame (i.e. substituting $r_k = k$ in equation 6-15) is equivalent to the sum of differences method described in section 2.4.2. In the absence of spatial speckle averaging (as described in the preceding section) the intermediate phase values cancel out and have no influence on the final unwrapped phase (other than to determine the integer valued correcting field $\nu$), thus no benefit is gained from integrating incremental unwrapped phase values in this way. However, when combined with spatial speckle averaging, the intermediate phase values no longer cancel out, contributing to the wrapped phase estimator in a maximum likelihood sense, enabling the maximum allowed deformation to be many times that for which total decorrelation would otherwise have occurred.

The practical benefits of summing incremental unwrapped phase maps in this way must be balanced against the significant bias that can be introduced in the final
unwrapped phase result when the re-referencing interval becomes small[97]. Frequent
re-referencing has the effect of integrating the errors present in the intermediate phase
values into the final unwrapped phase thus compromising one of the key benefits of
the modified sum of difference method described in section 2.4.3. Determination of
the optimal re-referencing rate for any given experimental arrangement continues to
be an area of active research at Loughborough University.

6.2.3 Temporal least squares fitting

Digital speckle pattern interferometry is often used to measure deformation of a test
surface under load (e.g. pressure, temperature, etc.). In general, the response of the
test surface, and therefore the unwrapped phase-change, can be expressed as a
function of the loading function, $P(t)$:

$$\Delta \phi(t,0) = f(P(t))$$ (6-19)

In practice, $P(t)$ and $f(P)$ are often reasonably continuous functions, allowing
equation 6-19 to be expanded as an $n$-order polynomial in $t$,

$$\Delta \phi(t,0) = p_0 + p_1 t + p_2 t^2 + \ldots + p_n t^n$$ (6-20)

In the presence of noise, the unwrapped phase change can be estimated by fitting a
curve as defined by equation 6-20 to the measured data in the least-squares sense. We
therefore aim to minimise the sum of squares of the error,

$$S = \sum_{t'=0}^{t} \left[ \Delta \phi(t',0) - \Delta \tilde{\phi}(t',0) \right]^2$$ (6-21)

We consider now only the case of a simple linear approximation to the surface
response, $\Delta \phi(t,0) = p_1 t$, and thus try to solve the partial derivative constraint,

$$\frac{\partial S}{\partial p_1} = 2 \sum_{t'=0}^{t} \left[ -\Delta \tilde{\phi}(t',0)t' + p_1 t'^2 \right] = 0$$ (6-22)

In this case, the least-squares estimator for the unwrapped phase is given by,

$$\Delta \tilde{\phi}(t,0) = \left[ \frac{\sum_{t'=0}^{t} \Delta \tilde{\phi}(t',0)t'}{\sum_{t'=0}^{t} (t')^2} \right]$$ (6-23)

In practice, equation 6-23 can be computed more efficiently by maintaining a running
weighted sum of the calculated unwrapped phases.
An instrument for real-time digital speckle pattern interferometry

\[ \Delta \tilde{\phi}(t, 0) = \Omega \sum_{t' = 0}^{t} [\Delta \hat{\phi}(t', 0)t'] \]  

(6-24)

where the scaling coefficient can be pre-calculated as,

\[ \Omega = \frac{t}{\sum_{t' = 0}^{t} (t')^2} \]  

(6-25)

The temporal least squares estimator can be combined with the re-referencing scheme described by equation 6-15, and is equivalent to fitting a straight-line to the incremental unwrapped phase data measured during each re-referencing interval independently. The total unwrapped phase-change is obtained as:

\[ \Delta \hat{\phi}(t, 0) = \Omega(t, r_k) \sum_{t' = r_k}^{t} \Delta \hat{\phi}(t', r_k)(t' - r_k) + \sum_{k=2}^{K} \left[ \Omega(r_k, r_{k-1}) \sum_{t' = r_{k-1}}^{r_k} [\Delta \hat{\phi}(t', r_{k-1})(t' - r_{k-1})] \right] + \Omega(r_1, 0) \sum_{t' = 0}^{t_1} [\Delta \hat{\phi}(t', 0)t'] \]  

(6-26)

where

\[ \Omega(t_1, t_2) = \frac{t_1 - t_2}{\sum_{t' = t_1}^{t_2} [(t' - t_2)^2]} \]  

(6-27)

Efficient implementations are made possible by pre-calculating the weighting coefficients using equation 6-27. Furthermore, if the re-referencing interval is constant such that \( r_k = kR \), then equation 6-26 simplifies to:

\[ \Delta \hat{\phi}(t, 0) = \Omega(t, r_k) \sum_{t' = r_k}^{t} \Delta \hat{\phi}(t', r_k)(t' - r_k) + \Omega(R, 0) \left\{ \sum_{k=2}^{K} \sum_{t' = r_{k-1}}^{r_k} [\Delta \hat{\phi}(t', r_{k-1})(t' - r_{k-1})] + \sum_{t' = 0}^{t_1} [\Delta \hat{\phi}(t', 0)t'] \right\} \]  

(6-28)

### 6.2.4 Simulation of extended temporal phase unwrapping algorithm

The issues discussed in sections 6.2.1 to 6.2.3 are best understood by way of an example. The speckle data recorded using a high-speed out-of-plane interferometer[97] and examined in section 3.4 is re-analysed using the extended TPUA algorithm.
Three methods of improving phase-change estimates by combining independent measurements are examined, and the resulting final unwrapped phase values compared to those obtained without combining measurements. The estimation methods are denoted as PE1 through PE4: PE1 does not spatially combine measurements; PE2 spatially averages the phase-change values by convolving the $\Delta \phi_w(t_1, t_2)$ images with a $3 \times 3$ pixel square filter function where the weighting of each sample is equal; PE3 is identical to PE2 except that the weighting of each sample is proportional to the product of the initial and final modulation (equation 6-13); and PE4 spatially averages the phase-change values by convolving the $F(t_1, t_2)$ and $G(t_1, t_2)$ images with a $3 \times 3$ pixel square filter function where the weighting of each sample is equal (equation 6-9).

We consider first the case where the reference speckle pattern is updated every processing frame in an attempt to reduce the problems of speckle decorrelation during the deformation process. Figure 6-2 shows the final unwrapped phase along row 112 (i.e. 80 rows from the bottom of the image) after 100 processing frames, computed using each of the phase estimation methods described above. Well-modulated pixels are unwrapped successfully using PE1, with intermediate phase values contributing only to the calculation of the integer valued correcting field, and can thus be shown to closely match the unwrapped phase result obtained by spatial phase unwrapping[97]. However, poorly modulated pixels unwrapped using PE1 exhibit significant unwrapping errors.

Phase-change estimator PE2 demonstrates a general improvement in the unwrapped phase calculated at poorly modulated pixels, however, errors are introduced at well-modulated pixels due to combining phase-change estimates from poorly modulated pixels with non-optimal weightings.

The unwrapped phase obtained using phase-change estimator PE3 shows a significant overall improvement in performance compared to PE2 demonstrating that more accurate phase-change estimates are obtained when using the theoretically optimal sample weightings.
Similarly, the phase-change estimator PE4 shows significant resilience at poorly modulating pixels, and results in an unwrapped phase at well-modulated pixels that closely matches that obtained using PE1.

Errors in the intermediate phase-change values no longer cancel out when using phase-change estimators PE2, PE3, and PE4, resulting in small phase errors present in each processing frame being integrated into the final unwrapped phase. These can be observed as positive or negative bias in the unwrapped phase graphs when compared to that obtained using PE1. Note, however, that the bias observed in reference [97] (figure 6) is larger than that of PE4 in figure 6-2 as the re-referencing interval used was reduced to one camera frame, rather than one processing frame (i.e. four camera frames) as used here.

Calculating phase-change values with respect to a single reference phase recorded at time \( t = 0 \) can reduce the accumulation of small phase errors, however, errors due to speckle decorrelation now become apparent. Figure 6-3 shows the final unwrapped phase computed using each of the phase estimation methods without updating the reference phase map. As expected, the unwrapped phase computed using phase-change estimator PE1 remains almost unchanged, as intermediate phase values cancel out.

Phase estimation by methods PE2 and PE3 results in significant errors in the unwrapped phase. This can be attributed to spatially combining phase-change values \( \Delta \hat{\phi}_{i,j}(t,0) \) that lie in the range \([-2\pi, 2\pi]\). The statistical benefit of spatially combining samples in this way is based on the assumption that speckle phase-change values vary slowly from pixel to pixel, however, when calculating the phase-change over long time intervals the resulting values can become wrapped. This clearly violates the above assumption and results in significant error in the phase-change estimate. This behaviour is avoided when re-referencing every frame, since the phase-change between consecutive phase values must lie in the range \([-\pi, \pi]\) if the data has been sampled in accordance with Shannon sampling theorem (as required for successful temporal phase unwrapping).
Phase-change estimation by method PE4, however, does not suffer from wrapped phase-change values since the estimate is computed by spatially combining values that are proportional to $\sin \Delta \hat{\phi}_w(t,0)$ and $\cos \Delta \hat{\phi}_w(t,0)$. The resulting unwrapped phase no longer suffers from the bias attributed to accumulation of small phase errors.

A compromise is sought in choosing a re-referencing interval that enables the benefits of these two scenarios to be combined. Figure 6-4 shows the results obtained when the reference phase map is updated every 10 processing frames. At well modulating pixels the unwrapped phase obtained using estimator PE4 continues to show good agreement with the results obtained using PE1, demonstrating that the accumulation of small phase errors from each processing frame is not significant. Furthermore, updating the reference phase map at regular intervals helps reduce errors due to speckle decorrelation introduced by deformation of the test surface. In practical situations, where periodic updates of the reference phase map are required to suppress speckle decorrelation-induced phase errors, spatial speckle averaging can be implemented using phase-change estimator PE4. However, phase-change estimators PE2 and PE3 should not be used together with re-referencing schemes as significant errors can be introduced when wrapped phase-change samples are combined.

We now take into account temporal least squares fitting to the observed unwrapped phase as described in section 6.2.3. We consider only the case of fitting a straight-line to the unwrapped phase values computed using estimator PE4 as implemented in the real-time system. Simulated results are obtained using the same experimental data as above since the non-linear specimen loading used in this experiment demonstrates the algorithm functionality. However, in practice, temporal least squares fitting should not be used in this situation, and does not result in improved signal to noise ratio.

Figure 6-5 shows the final unwrapped phase computed along row 112 after 100 processing frames both with and without least-squares fitting along the time axis. Trace PE4 is obtained when the reference phase map is not updated during the phase evaluation and no least-squares fitting has been applied (this corresponds to trace PE4 in figure 6-3). Trace PE4:LSF is the result of modifying the phase estimate by fitting a straight-line to the data at each pixel in a least-squares sense. However, the modified unwrapped phase estimate falls significantly below that obtained without fitting. This
An instrument for real-time digital speckle pattern interferometry can be understood by examining figure 6-6, which plots unwrapped phase calculated for a single pixel at (100,112) against the processing frame index. Trace PE4 clearly shows that the surface displacement is non-linear with respect to time, and therefore attempts to fit a single straight-line segment to the observed data results in large errors in the modified estimate.

Trace PE4: REF10 is obtained when the reference phase map is updated every 10 processing frames and no least-squares fitting has been applied (this corresponds to trace PE4 in figure 6-4). Trace PE4: REF10: LSF uses the same re-referencing scheme and applies temporal least-squares fitting along the time axis. In this case the algorithm attempts to fit independent straight-line segments to the unwrapped phase observed in each re-referencing interval. The phase gradient of the first straight-line segment is computed using only the first 10 consecutive unwrapped phase estimates obtained using the initial reference phase map. An independent gradient estimate for the second line segment is then computed using the following 10 consecutive unwrapped phase estimates. As this procedure continues, the resulting unwrapped phase estimate shown in figure 6-6 can be seen to follow the non-linear surface response much more closely, however, the final unwrapped phase estimate is still somewhat low.

In conclusion, straight-line fitting to unwrapped phase data along the time axis in a least-squares sense is not beneficial if the surface response is non-linear with respect to time. However, simulation of the least squares fitting algorithm using non-linear experimental data demonstrates the algorithm behaviour and the effect of the re-referencing parameter (and therefore the fitting interval) more clearly than can be achieved using linear data.

6.3 Mapping extended algorithm to pipeline image processing elements
A real-time DSPI system has been implemented based on the pipeline image processor platform, and includes the functionality of the extended temporal phase unwrapping algorithm described in section 6.2. The implementation includes the ability to perform spatial speckle averaging using phase-change estimator PE4, to periodically update the reference phase map, and to improve the phase estimate using
An instrument for real-time digital speckle pattern interferometry

temporal least-squares fitting. The system supports acquisition of phase-modulated
speckle interferograms at 60 frames s\(^{-1}\) and displays pseudo-colour unwrapped phase
maps at 15 frames s\(^{-1}\). The following paragraphs describe mapping of the algorithm to
the limited hardware resources available on the two MaxVideo250 pipeline image
processing boards described in section 3.5, and make reference to the simplified
schematic diagrams denoted D1 through D6 included in appendix A8. The standard
four-frame phase-stepping algorithm \((N = M = 4)\) has been chosen for the DSPI
application enabling some of the pipeline configurations implemented for the real-
time profilometry system to be re-applied. In particular, image acquisition and
calculation of the \(N(t)\) and \(D(t)\) images remains as described in section 4.5.3.
Synchronisation of the reference beam phase modulator to the image acquisition
sequence is handled by the software application and is discussed further in section
6.5.

Under normal operation the sequence of hardware arrangements denoted D1 through
D4 is applied once for each processing frame. In contrast to the profilometry
implementation where the number of processing frames can be determined before
starting the measurement based on the maximum fringe density and the exponential
coefficient, the DSPI implementation continues cycling through the sequence of
hardware arrangements until the user indicates otherwise. This may be specified
explicitly by defining the measurement duration, or simply by pressing a key on the
computer keyboard to terminate the measurement. The user is able to choose whether
to display a greyscale wrapped phase map or a pseudo-colour unwrapped phase map,
which is updated once for each processing frame (i.e. 15 frames s\(^{-1}\)).

Hardware configuration D1 includes an image acquisition pipeline to capture
interferogram \(I(0)\) and a continuous display pipeline that feeds the analogue video
signal generator (DAC1) used to drive the SVGA display monitor. Image surfaces
\(N(t)_{B_0}\) and \(D(t)_{B_0}\) store the \(N(t)\) and \(D(t)\) images, respectively, calculated by
hardware configurations D3 and D4 during the previous processing frame. Surfaces
\(N(r_k)_{B_0}\) and \(D(r_k)_{B_0}\) store the \(N(r_k)\) and \(D(r_k)\) images, respectively, that were
calculated from the most recent reference frame (e.g. 0, \(r_1\), \(r_2\), ...). The remaining
pipeline calculates the \(F(t,r_k)\) image within the linear arithmetic unit \(AL_0\) and
An instrument for real-time digital speckle pattern interferometry performs spatial speckle averaging in the neighbourhood multiply and accumulate element (NMAC) as described by the numerator of equation 6-9. Lookup table AB_LUTO is required to negate the $N(r_k)$ data stream in order to implement the subtraction operator in equation 2-50. The hardware NMAC operator supports square kernels up to $8 \times 8$ pixels with 12-bit coefficients, and takes an input stream with a maximum word length of 12 bits. The coefficient implemented at the shift element within the arithmetic unit scales the $F(t,r_k)$ data stream and must be chosen carefully to avoid saturating the NMAC input, whilst at the same time avoiding the loss of significant bits. A further shift element is used to scale the NMAC result such that it does not exceed the dynamic range of the 16-bit output path. The 16-bit result, denoted $F'(t,r_k)$, is then stored in the pair of image surfaces denoted as $F'(t)_{B1:B0}$.

The acquisition and display pipelines included in hardware configuration D2 are identical to those in D1. The receive gateway to $I_0(t)_{B0}$ is used to double buffer image memory such that interferogram $I(1)$ does not overwrite $I(0)$. Linear arithmetic unit AU_L0 is now used to calculate $G(t,r_k)$ as described by equation 2-51, which is then convolved with a square filter function as described by the denominator of equation 6-9, and scaled to produce the 16-bit result $G'(t,r_k)$. In general, speckle interferograms exhibit large variations in modulation from pixel to pixel and hence it is preferable to normalise real and imaginary components of the complex amplitude before calculating the argument when using a lookup table with fixed-point number representation (see the discussion regarding TPUA8 in section 3.4). The most significant 8 bits of $F'(t,r_k)$ and $G'(t,r_k)$ are therefore used as operands to the AP_WORD lookup table which stores pre-calculated normalising coefficients equal to $1/\text{MAX}(|F'(t,r_k)|,|G'(t,r_k)|)$, where the MAX() function returns the larger operand. The 16-bit coefficient returned by the lookup table is stored in the pair of image surfaces denoted as $M(t,r_k)_{B1:B0}$, and passed as a scaling operand to linear arithmetic unit AU_L1 where $G'(t,r_k)$ is normalised to 8 bits. The resulting 8-bit data stream $G''(t,r_k)$, proportional to $\cos \Delta \tilde{\phi}_n(t,r_k)$, is then stored in the pair of image surfaces denoted $G''(t)_{B1:B0}$. Variable J1 selects which operand is passed to the output of the multiplex elements, and can be used to update the reference images stored in surfaces.
An instrument for real-time digital speckle pattern interferometry

$N(r_k)_{B0}$ and $D(r_k)_{B0}$. Depending on the value of $J1$, the multiplex elements will either pass the existing reference data (i.e. no change in the reference phase data), or select the data from the current processing frame (i.e. perform re-referencing of the phase data).

The acquisition pipeline included in hardware configuration D3 computes the $N(t)$ image by subtracting the interferogram $I(2)$ transmitted by the camera from interferogram $I(0)$, and stores the result in image surface $N(t)_{B0}$. The display pipeline continues to feed image data to the analogue video generator. The remaining pipeline utilises most of the available hardware to calculate a new wrapped phase-change estimate $\Delta \phi_n(t, r_k)$ and update the integer valued correcting field $\nu(t, r_k)$. Linear arithmetic unit AU_L1 scales the $F'(t, r_k)$ using the normalising coefficients stored in $M(t, r_k)_{B1:B0}$ to produce an 8-bit data stream proportional to $\sin \Delta \phi_n(t, r_k)$. Denoted $F''(t, r_k)$. The two 8-bit data streams, $F''(t, r_k)$ and $G''(t, r_k)$, are combined to form a 16-bit operand to lookup table AU_L1, which stores pre-calculated results for the four-quadrant arctangent function. The resulting 16-bit phase-change value, $\Delta \phi_n(t, r_k)$, is stored in the pair of image surfaces denoted $\Delta \phi_n(t, r_k)_{B1:B0}$. Non-linear arithmetic unit AU_N1 subtracts the lookup table result obtained during the previous processing frame from the current result to generate $\Delta \hat{\phi}_n(t, r_k) - \Delta \phi_n(t-1, r_k)$. Non-linear arithmetic unit AU_N0 implements the NINT() function of equation 6-18 to produce $d(t, r_k)$, which is used to update the $\nu(t, r_k)$ according to equation 6-17.

Image $D(t)$ is computed in hardware configuration D4 using the acquisition pipeline to subtract interferogram $I(1)$ from interferogram $I(3)$ transmitted by the camera. The result is stored in image surface $D(t)_{B0}$. The display pipeline continues to feed image data to the analogue video generator. The remaining hardware is used to compute the new unwrapped phase map $\Delta \hat{\phi}(t, 0)$ and re-calculate the display data. Unwrapped phase-change $\Delta \hat{\phi}(t, r_k)$ is calculated by non-linear arithmetic unit AU_N1 following equation 6-16, where multiplication of the $\nu(t, r_k)$ operand by constant $2\pi$ is achieved by simply bit-shifting the data with respect to the $\Delta \phi_n(t, r_k)$ operand. In order to provide sufficient dynamic range for the unwrapped phase data without
discarding significant bits from the intermediate results, the $\Delta \phi(t,0)$ values are stored as 24-bit values across three image surfaces denoted by $\Delta \phi(t,0)_{B2,B1,B0}$. Unfortunately, however, the arithmetic processors provided in hardware do not directly support operations on 24-bit operands. Therefore, implementation of equation 6-16 is somewhat complicated by the need to split the addition operation across both the linear- and non-linear arithmetic units, AU_L0 and AU_N0, respectively. Bit carry/borrow between the two halves of the operation is implemented using the multiplex element in AU_N0. Input J8 is used to program the display mode by controlling the multiplexer elements that select the operands to the linear arithmetic unit AU_L1. Depending on the operands selected, the display can represent: (a) wrapped phase-change with respect to the last reference phase, $\Delta \phi(t, r_k)$; (b) unwrapped phase-change with respect to the last reference phase, $\Delta \phi(t, r_k)$; or (c) unwrapped phase-change with respect to the initial phase, $\Delta \phi(t,0)$. Linear arithmetic unit AU_L1 scales the display data according to constants J9 and J10, then adds a bias specified by J11. Finally, the lookup table AP_WORD applies the colourmap conversion to generate pseudo-colour image data before writing the result to the display image surfaces denoted DispB1:B0.

The real-time DSPI system supports a limited implementation of temporal least-squares fitting combined with regular updates of the reference phase map. The least-squares estimator is based on a linear approximation, and is evaluated using equation 6-28, with the exception that the first term has been omitted from the implementation. This introduces some constraints when using the temporal least-squares fitting feature: (a) the re-referencing interval must be constant and known before starting the measurement to allow pre-calculation of the weighting values (using equation 6-27); and (b) measurement must terminate after an integer multiple of the re-referencing interval (otherwise the first term in equation 6-28 would take a non-zero value, resulting in significant phase errors). The additional computation needed to evaluate the estimator requires a modified sequence of hardware configurations. When the user enables temporal least squares fitting the hardware sequence becomes D1, D2, D3, D5, D6, where data received from the camera during configuration D6 is discarded.
An instrument for real-time digital speckle pattern interferometry

and the state of the reference beam phase modulator is not updated. The additional computation time results in a reduced display update rate of 12 frames s\(^{-1}\).

Hardware configuration D5 includes the same acquisition and display pipelines as D4. Similarly, computation of the unwrapped phase-change \(\Delta \hat{\phi}(t, r_k)\) by non-linear arithmetic unit AU_N1 remains unchanged. The remainder of D5 is dedicated to computing the least-squares phase-change estimator according to the second term in equation 6-28. Input J12 applies a pre-calculated weighting to \(\Delta \hat{\phi}(t, r_k)\) before re-calculating the summation operator using the linear and non-linear arithmetic units, AU_L0 and AU_N0, respectively. The 24-bit result is stored across the three surfaces denoted \(\Sigma(t, r_k)_{B2:B1:B0}\).

Hardware configuration D6 is very similar to configuration D4, with the exception that it does not include an image acquisition pipeline and so discards data received from the camera. The temporal least-squares estimator evaluated by D5 is used to recalculate the unwrapped phase image, \(\Delta \hat{\phi}(t,0)\) and update the display data.

6.4 Peripheral system components

6.4.1 General purpose interface bus subsystem

The real-time DSPI temporal phase unwrapping system communicates with peripheral hardware systems using an industry standard General Purpose Interface Bus (GPIB). A Greenspring IP488 GPIB controller module residing on a VIPC616 carrier board has been installed in the pipeline image processor VME backplane. The software driver distributed with the Greenspring GPIB controller was not compatible with the real-time LynxOS operating system and it was therefore necessary to specify a custom driver requirement for the development platform. Arcom Control Systems Ltd were contracted to port the existing software driver to the new specification. The GPIB interface enables the system to communicate with several devices in the laboratory, including an Instron 4466 load frame (fitted with a 10 kN load cell), a Gould 1602 dual channel digital oscilloscope, and a Microlink multi-channel signal input / output interface unit. Direct control of the Instron load frame closes the measurement loop, and in principle (although not implemented here) enables the specimen loading to be
varied in response to the measured phase data. Figure 6-7 provides a simplified schematic overview of the system configuration.

6.4.2 Phase modulators

The Microlink interface unit includes four digital to analogue converters (DACs) that can be used to drive a variety of phase modulators (see figure 6-8), and has been successfully demonstrated with: (a) a glass wedge mounted on a Physik Instruments P820-10 low-voltage PZT, with a nominal extension characteristic of 0.15 µm V⁻¹; (b) a mirror mounted on a Burleigh PZ-81 high-voltage PZT, with a nominal extension characteristic of 2.0 nm V⁻¹; and (c) a Leysop high-voltage Pockels cell, which introduces a nominal phase change of 17.7 mrad V⁻¹ (equivalent to an optical path length change of 1.5 nm V⁻¹ for λ=532 nm). In the case of the latter two modulators, the analogue drive voltage is amplified using device specific high-voltage amplifiers, designed for driving predominantly capacitive impedance loads.

The ability to select from a range of phase modulators enables the system to be adapted to a variety of interferometer arrangements. Initial experiments were conducted using a flexible interferometer design, with two single-mode, polarisation-preserving optical fibres delivering the object and reference beams. The phase modulator is located between the laser and the reference beam fibre; either an in-plane or an out-of-plane interferometer arrangement can be configured by suitable positioning of the output ends of the fibres, with no further change in optical set-up required. Figure 6-9 shows a photograph of the interferometer in an out-of-plane arrangement, with an annotated schematic of the set-up included in figure 6-10 (see appendix A9 for detailed drawings of the Microbench™ arrangements). The light source is a 200 mW frequency doubled Nd:YAG laser (532 nm) manufactured by Lightwave. Phase-stepping is carried out by means of a glass wedge mounted on a Physik Instruments P820-10 low-voltage PZT[94] as shown schematically in figure 6-11. This phase-stepping method has several advantages over the conventional approach of reflecting the beam from a PZT-mounted mirror: (a) no beam tilt is produced if the PZT tilts during translation, and (b) extension of the PZT does not introduce lateral beam translation provided the wedge is angled correctly in the beam. The complete phase stepping assembly and fibre optic launch optics are included in a
An instrument for real-time digital speckle pattern interferometry

A compact unit constructed using Spindler and Hoyer Microbench™ components and Point Source fibre optic mountings (figure 6-12).

The Burleigh PZ-81 high-voltage PZT is particularly suited to Michelson interferometer arrangements, since the end face of the transducer includes a screw thread for mounting small circular mirrors. The device includes three independently controllable piezoelectric stacks, allowing the extension and tilt to be manipulated under computer control. In phase-shifting applications parallel extension of each of the stacks enhances the stability of the translated mirror compared to devices employing only a single piezoelectric stack. The transducer has been successfully used in shearing interferometer arrangements of the form shown in figure 1-7, where one of the mirrors is mounted on the end of the device and translated perpendicularly to the reflecting surface to introduce a path length difference between the interferometer arms. The other mirror is typically mounted on a three-axes adjustable mount, and tilted slightly to introduce the image shear.

The Leysop Pockels cell is able to support much higher phase-stepping rates than PZT actuators, as there are no mechanical moving parts. Crystal birefringence is controlled by an applied electric field, enabling the retardance to be varied as desired. Whilst the real-time DSPI system has a constant phase-stepping time interval of 1/60 s, optical arrangements are often shared with systems that use cameras with much higher framing rates (e.g. 1,000 frames s⁻¹) for which PZT actuators are unsuitable.

The phase modulators described in this section have been calibrated following the procedure described by Ochoa et al.[155]. The procedure is only suitable for calibrating phase modulators that can produce a full $2\pi$ phase shift, and for which the phase changes monotonically with the applied voltage, but is relatively simple to implement and can be performed in situ. The technique can be applied by interfering either speckle or smooth wavefronts, and is based on evaluation of the spatial mean square difference, $S$, between reference intensity $I_0$ and the intensity $I$ obtained after an unknown phase step introduced in one of the interferometer arms:

$$S = \left\langle (I - I_0)^2 \right\rangle$$
The quantity $S$ varies sinusoidally as a function of the applied voltage change $(V - V_0)$, and can be approximated by the function,

$$S(V, V_0) = A + B \cos[\Delta \phi(V, V_0)].$$

(6-30)

where $A$ and $B$ are constant, and $\Delta \phi(V, V_0) = \phi(V) - \phi(V_0)$ is a slowly varying function describing the unknown phase-change as a function of the applied voltage and the reference voltage used for $I_0$. The phase-change function can in turn be approximated by,

$$\Delta \phi(V, V_0) = a(V - V_0) + b(V^2 - V_0^2),$$

(6-31)

where $a$ and $b$, the linear and non-linear calibration coefficients, respectively, can be estimated by fitting $S(V, V_0)$ to the measured data in a least-squares sense.

The phase modulators described above were calibrated using a reference voltage $V_0 = 0$, and in each case the non-linear calibration coefficient was found to be negligible. The linear calibration coefficient for the Physik Instruments, Burleigh, and Leysop phase modulators were estimated as $4.8 \times 10^{-1}$ rad V$^{-1}$, $4.6 \times 10^{-2}$ rad V$^{-1}$, and $1.7 \times 10^{-2}$ rad V$^{-1}$, respectively.

6.5 Synchronisation

Significant errors in the estimated phase can occur in phase-stepping interferometers when the photodetector and phase modulator are not synchronised correctly. Variations in the phase modulator output during the integration period of the photodetector results in reduced signal modulation as discussed in section 2.2. A trace of the phase modulator output should therefore ideally follow a “staircase” appearance, where steps are synchronised to the photodetector integration period, and occur at the inter-frame boundary. In practice, several factors can disturb this synchronisation scheme when using computer controlled phase modulators: (a) problems associated with identification of the photodetector inter-frame boundary; (b) errors in the user-level software application; (c) delay introduced by the operating system kernel-level software drivers; (d) delay introduced by the hardware interface between the computer and the phase modulator; and (e) poor actuator response to the
An instrument for real-time digital speckle pattern interferometry

command signal (e.g. low-frequency mechanical response characteristic, overshoot, and oscillation).

The approach taken for the phase-stepping control used in both the real-time profilometry and DSPI systems is based on synchronisation to a master timing signal, which is derived from the digital camera output. The EEV CAM17 camera transmits a 20 MHz pixel clock that regulates the flow of image data through the pipeline processor hardware, and horizontal and vertical synchronisation signals (denoted as Hsync and Vsync, respectively). The phase-stepping software is able to synchronise to these signals and ensure that commands issued to the phase modulator occur at the photodetector inter-frame boundary. In the case of the real-time DSPI application, this involves transmitting GPIB commands to the Microlink unit, where a DAC generates an analogue drive signal for the phase modulator.

The phase modulators used with the real-time DSPI system were studied to evaluate their performance under software control, and to validate that their response was synchronised to the photodetector inter-frame boundary. For brevity, the following discussion considers only the Burleigh PZ-81 PZT, presenting the experimental set-up and results.

A Michelson interferometer arrangement was constructed from Spindler and Hoyer Microbench™ components to form the interference pattern from two parallel smooth wavefronts (see appendix A10), which was recorded using a Kodak Ektapro high-speed camera. The Lightwave 200 mW frequency-doubled Nd:YAG laser was used as a light source, using a single-mode, polarisation-preserving optical fibre to deliver the beam. A positive achromatic lens was used to form a collimated beam, which was then divided using a beam-splitter cube. One of the mirrors was mounted on the computer controlled Burleigh PZT, and the other on a 3-axis optical mount. The tilt angle between the interferometer arms was adjusted using the optical mount to generate approximately eight parallel vertical fringes across the CCD element. Extension of the PZT introduces an optical path length change in one of the interferometer arms, resulting in a uniform phase shift in the fringe pattern, which is recorded by the camera. Figure 6-13 shows a subset of four frames taken from an image sequence recorded by the Ektapro camera at 6,000 frames s⁻¹, the phase
An instrument for real-time digital speckle pattern interferometry modulator was controlled by the real-time DSPI application and is therefore synchronised to the EEV CAM17 camera running at 60 frames s$^{-1}$.

The relative position of the PZT-mounted mirror can be determined by post-processing the image data to extract the phase information. Each row in the image was analysed using the one-dimensional fast Fourier transform, zero padding the data to increase the frequency resolution, and the phase determined from the real and imaginary coefficients for the dominant frequency. The average phase over all rows within the image was then calculated. A Gould 1602 dual channel digital oscilloscope was used to record the Vsync synchronisation signal transmitted by the CAM17 camera and the analogue PZT drive voltage. The resulting traces are shown in figure 6-14, together with the extracted phase information. The PZT drive voltage clearly shows the "staircase" appearance corresponding to the four-frame phase stepping algorithm ($M = N = 4$), with the voltage transitions synchronised to the camera Vsync pulses. The resulting phase modulation follows the general form of the drive signal, however, the step rise time is extended due to the mechanical response of the actuator. Small oscillations on the phase response are likely to be attributed to mechanical overshoot and the corresponding ringing effect; the amplitude of the oscillations can be seen to fall away over the camera frame interval. Over several such experiments, it was noted that the response time (rise time) of the mechanical actuator was consistently very short compared to the photodetector integration period, and that the amplitude of the mirror oscillations was small compared to the step size. However, it was also noted that individual step transitions in the PZT drive voltage (and consequently the phase) could sometimes be delayed for up to several milliseconds compared to those shown in figure 6-14, and that these delays appeared to occur at random.

It was initially suspected that these delays could be attributed to software issues, either caused by the operating system kernel suspending the DSPI application in favour of another software task, or perhaps a delay introduced by the kernel GPIB driver. The software priority of the DSPI application was therefore raised above other tasks running on the system, and further measurements were taken to observe the system behaviour close to the step transition.
The DSPI application was modified slightly to enable the time at which the user-level application issues the GPIB commands to the kernel driver to be recorded. In the absence of a suitable hardware timer that could be accessed by the software application, the approach taken here was to toggle the state of one of the signal lines in the RS232 serial interface connector, which could then be monitored by the digital oscilloscope. However, it was found that the data buffering scheme employed by the kernel RS232 driver and hardware introduces an unacceptable delay between a user-level software data transmit request and the corresponding change in signal level at the connector. Fortunately, the RS232 standard includes a simple “handshaking” signal denoted as DTR (data terminal ready) that indicates that the device is online. This signal is not buffered either by the kernel driver or in hardware, and enables very fast switching with little overhead. The DSPI application was therefore modified to toggle the state of the DTR signal immediately before, and immediately after, issuing GPIB commands to the kernel driver.

The GPIB bus employs an 8-bit parallel protocol, where valid data is indicated by the state of a “data available” signal (DAV). Recording the DAV signal level using an oscilloscope therefore enables activity on the bus to be monitored. However, the Gould 1602 oscilloscope is only able to record two channels simultaneously, and therefore the recordings could not be made whilst also capturing the Vsync and RS232 DTR signals. The GPIB activity was therefore captured in a separate experiment, using the remaining channel to capture the camera Vsync signal and thus allowing the data to be merged at a later time. A similar approach was taken to record the analogue PZT drive voltage from the Microlink DAC output.

Figure 6-15 shows a typical result, created by merging the oscilloscope trace data, where the relative temporal position of each trace has been calculated with respect to the rising edge of the Vsync signal. The user-level DSPI application issues the GPIB commands during the DTR pulse interval; there is a delay of approximately 0.15 ms before a burst of activity is observed on the GPIB bus; the output of the DAC is updated approximately 0.6 ms after the user-level application issued the GPIB commands (equivalent to 4% of the photodetector integration period). Some GPIB activity can be seen following the DAC output transition and this is caused by the...
An instrument for real-time digital speckle pattern interferometry

DSPI application pre-configuring the DAC ready to receive the output commands at the next inter-frame boundary.

It was noted that random delays in the PZT drive signal transitions with respect to the Vsync signal were not accompanied by a corresponding delay in the DTR transitions, but appeared to be associated with an extended burst of activity on the GPIB bus. Investigation revealed that the bus relies on remote equipment acknowledging valid receipt of data, and when this is not received, the data is re-transmitted. The observed random delays were therefore attributed to the effects of electrical interference on the data transmission path, and the corresponding re-transmission time. The plastic GPIB connectors and standard cabling supplied with the Greenspring GPIB module were therefore replaced with metal connectors and industrial grade shielded cable. Further experiments with the new set-up revealed that random delays in the PZT drive signal had been eradicated, and that the timing information shown in figure 6-15 could be achieved repeatably.

6.6 Summary

Whole-field DSPI techniques enable the measurement of rough surface deformation with high sensitivity, however, such systems are susceptible to the effects of air turbulence and vibration noise. Video-rate acquisition systems reduce measurement errors from these noise sources, and enable the continuous deformation of objects under time-varying loads to be measured. A DSPI system has been developed that enables real-time quantitative analysis of speckle interferograms, and complements the real-time profilometry system described in chapter 4. The DSPI system and profilometry system share the same pipeline image processing platform and are controlled by a common software application. Measurements are made using the same camera allowing direct spatial registration of results from each method. The DSPI system is based on the standard four-frame phase-stepping algorithm, and acquires phase-stepped interferograms at 60 frames s\(^{-1}\). Greyscale wrapped phase maps or pseudo-colour unwrapped phase maps are displayed in real-time at 15 frames s\(^{-1}\). The system is compatible with a variety of phase modulators, including a Physik Instruments low-voltage PZT, a Burleigh high-voltage PZT, and a Leysop Pockels cell.
An instrument for real-time digital speckle pattern interferometry

The system implements an extended form of the temporal phase unwrapping algorithm that is designed to improve signal to noise in phase estimates from speckle interferograms, and includes support for: spatial speckle averaging over an area of up to 8 x 8 pixels; periodic updates of the reference phase map; and first order temporal least-squares fitting to the unwrapped phase (at a reduced display rate of 12 frames s$^{-1}$).

A high-level description of algorithm mapping to the pipeline image processor hardware was presented and the phase modulator synchronisation scheme described. Random delays observed in the phase modulator output transitions were studied, and attributed to electrical interference on the GPIB bus. Replacement cabling and careful shielding of the electrical signals has enabled repeatable, deterministic phase modulator timings to be obtained.
6.7 Figures

Figure 6-1: Phasor diagram of the complex amplitudes measured at a single pixel in the image plane. Vectors $a_0(t_1)$ and $a_0(t_2)$ correspond to the complex amplitudes measured by a phase stepping interferometer observed at times $t_1$ and $t_2$, respectively. The angle between them is the observed phase change $\Delta \hat{\phi}$, which differs from the angle of interest $\Delta \phi$ due to speckle decorrelation.

Figure 6-2: Cross section along row 112 of the unwrapped phase difference map $\Delta \phi(t,0)$ computed after 100 processing frames and obtained by updating the reference phase map every processing frame. Results are presented for each of the four phase-change estimators described in section 6.2.4.
Figure 6-3: Cross section along row 112 of the unwrapped phase difference map $\Delta \phi(t,0)$ computed after 100 processing frames and obtained without updating the reference phase map. Results are presented for each of the four phase-change estimators described in section 6.2.4.
Figure 6-4: Cross section along row 112 of the unwrapped phase difference map $\Delta \phi(t,0)$ computed after 100 processing frames and obtained by updating the reference phase map every 10 processing frames. Results are presented for each of the four phase-change estimators described in section 6.2.4.
Figure 6-5: Cross section along row 112 of the unwrapped phase difference map $\Delta \phi(t,0)$ computed after 100 processing frames using phase-change estimator PE4. The plot compares the unmodified unwrapped phase estimates with those modified by fitting straight-line segments to the data at each pixel in a least squares sense. Trace PE4 and PE4:LSF were obtained without updating the reference phase map, and correspond to the unmodified and modified phase estimates, respectively. In trace PE4:LSF all unwrapped phase values at a given pixel contribute to the gradient estimate of a single straight-line segment. Trace PE4:REF10 and PE4:REF10:LSF were obtained by updating the reference phase map every 10 frames, and correspond to the unmodified and modified phase estimates, respectively. In trace PE4:REF10:LSF straight-line segments are fitted to the data in each referencing interval independently. Note that result PE4:REF10 is very similar to PE4, and almost obscures the latter trace.
Figure 6-6: Unwrapped phase difference estimate $\Delta \phi(t,0)$ computed for a single pixel at (100,112) plotted against the processing frame index (which is proportional to time). The unwrapping methods are the same as defined for figure 6-5. Fitting a single straight-line segment to the non-linear surface response results in significant errors in the modified phase estimate as can be seen in trace PE4:LSF. Fitting shorter independent straight-line segments enables the modified phase estimate to follow the surface response much more closely as can be seen in trace PE4:REF10:LSF. In general, the modified phase estimate will only demonstrate an improved signal to noise if the fitting interval, and therefore the length of the straight-line segments, is chosen with care. Significant errors can be introduced if these parameters are set inappropriately or if the physical quantity of interest is non-linear with respect to time. Note that result PE4:REF10 is very similar to PE4, and almost obscures the latter trace.
Figure 6-7: Schematic diagram showing peripheral components used in the real-time pipeline image processing system.

Figure 6-8: Photographs of the phase modulators compatible with the real-time DSPI system: (a) Physik Instruments P820-10 low-voltage PZT actuator. The small cylindrical device has a total length of 26 mm and diameter of 9 mm. The PZT is typically used to translate a small glass wedge in the phase stepping arrangement shown in figure 6-11; (b) Burleigh PZ-81 high-voltage PZT actuator. The cylindrical device is shown mounted in a custom-made rectangular aluminium housing that provides a simple mechanical interface to other optical components. The device is typically used in Michelson interferometer arrangements, where a small circular mirror (diameter 30 mm) is mounted on the end face; (c) Leysop Pockels cell. The cylindrical device has a length of 120 mm and consists of two ammonium dihydrogen phosphate (NH₄H₂PO₄, often referred to as ADP) crystals with an aperture of 3.5 mm. The two crystals are arranged such that electric field induced phase change is additive, but so that beam offset due to natural birefringence cancels out.
Figure 6-9: Photograph of 200 mW Nd:YAG Mach-Zehnder interferometer using single-mode, polarisation-preserving optical fibres to deliver object and reference beams in an out-of-plane arrangement.

Figure 6-10: Schematic of Mach-Zehnder interferometer shown in figure 6-9.
Figure 6-11: Schematic of phase modulator arrangement using a glass wedge mounted on Physik Instruments P820-10 low-voltage PZT. The polarised beam is divided into two by the polarising beam-splitter cube, and the two beams launched into single-mode, polarisation-preserving optical fibres. The split ratio between the two beams is controlled by rotation of the half-wave plate. Phase-stepping is performed by moving the PZT-mounted 45° glass wedge across one of the beams; a second identical wedge then brings the beam parallel to its original optical axis. The lateral displacement of the beam is constant and independent of the PZT extension provided the orientation of the glass wedges is chosen correctly. In the case of BK7 glass wedges ($n=1.519$), as used in the experimental arrangement discussed here, the prism (hypotenuse) surface should be orientated at 63.7° relative to the incident beam[156].
Figure 6-12: Photograph of Spindler and Hoyer Microbench™ implementation of the phase modulator arrangement shown in figure 6-11. The 3-axis optical mount can be seen in the foreground, with the PZT mount just visible in the background. The phase-stepped beam is launched into the armoured fibre optic cable seen on the right, and the other beam launched into the fibre optic at the top of the image.
An instrument for real-time digital speckle pattern interferometry

Figure 6-13: Subset of frames taken from a sequence of interferograms acquired using a Kodak Ektapro camera running at 6000 frames s\(^{-1}\). The fringes were formed from the interference of two smooth coherent wavefronts, where the phase of one of the wavefronts was modulated using a mirror mounted on a Burleigh PZ-81 PZT actuator. Image frames 050, 150, 250, and 350 shown here correspond to phase steps of 0, π/2, π, and 3π/2 radians that were applied with the period of 1/60 s.
Figure 6-14: The analogue output signal from the Microlink unit is synchronised to the vertical synchronisation signal (Vsync) received from the EEV CAM17 camera, and amplified by the Burleigh RG93 high voltage amplifier before being applied to the PZT actuator. The PZT drive signal was measured at the Microlink output and has a measured dynamic range of 2.1 V<sub>p-p</sub>. The position of the PZT has been determined by extracting the phase from a sequence of images of vertical fringes (recorded at 6000 frames s<sup>-1</sup>) modulated by a mirror mounted on the actuator. The measured fringe phase has a dynamic range of 4.7 rad, corresponding to phase steps of π/2 rad. The response time of the PZT actuator is very small compared to the integration period of the CAM17 camera. Furthermore, the amplitude of overshoot and subsequent oscillations observed in the actuator position are small compared to the step size.
Figure 6-15: Analysis of time delay between the start of the integration period of the CCD and the corresponding voltage step applied to the PZT phase modulator by the GPIB controlled DAC. The traces were recorded using a Gould 1602 dual channel digital oscilloscope in three separate experiments, where the Vsync signal was recorded on channel 1 in each experiment. Each trace has been plotted with respect to the rising edge in the Vsync signal that marks the start of the CCD integration. The red trace records the voltage on the DTR pin of the RS232 connector, and is toggled by the software application to mark the start and end of the section of code that passes GPIB commands to the operating system kernel. The blue trace records the voltage of the DAV pin of the GPIB connector that is used to signal that valid data is available on the GPIB bus. The 8-bit bus clearly introduces a finite delay as the multi-byte commands are transmitted to the receiving device. The receiving device in this case is a digital to analogue converter, and the change in output corresponding to the received commands can be seen in the green trace.
7 Application of the real-time digital speckle pattern interferometry instrument

7.1 Introduction

This chapter is concerned with experimental validation of the real-time temporal phase unwrapping system described in the preceding chapter and includes an example of application to defect detection in composite materials using digital speckle pattern interferometry.

The phase-stepping interferometer described in the preceding chapter is used to measure out-of-plane displacement of a flat plate displaced along the camera axis by a piezo-electric transducer driven by a periodic signal. Unwrapped phase values computed by the real-time system are presented.

A hardware fault that developed in the Lightwave laser prevented the validation experiments from being completed using this interferometer arrangement. A new phase-stepping interferometer arrangement is described based on a 5 W continuous wave laser, and is used to measure out-of-plane displacement of a flat plate rotated linearly with respect to time about an axis perpendicular to the camera axis.

The development of a vacuum hood suitable for loading of flat surfaces and small samples is described. Out-of-plane surface displacement of a carbon fibre sample due to small changes in environmental pressure is measured using the real-time DSPI system and vacuum hood. Experimental results are presented clearly showing the location of a complex artificial disbond below the surface.

Unfortunately, hardware faults encountered during the course of the work described in this thesis were not limited to those of the Lightwave laser. In particular, the CAM17 digital camera manufactured by EEV proved to be particularly unreliable, developing electronic faults on four separate occasions. In each case, it was necessary to ship the device to the manufacturer for repair, introducing a delay of several weeks. Problems have also been experienced with the Datacube pipeline image processor. During development of the real-time DSPI system it was noted that integer arithmetic results
Application of the real-time digital speckle pattern interferometry instrument computed by the Advanced Processor module located on one of the MAX250 VME boards was sometimes in error. Investigations by the author revealed that one of the arithmetic bits in the result was not being set correctly and this was traced to a poor electrical connection between the main printed circuit board and the plug-in module. Cleaning of the electrical connector and re-seating the module finally resolved the problem. More recently, one of the MV250 boards has developed a hardware fault associated with image memory requiring that the device be shipped to the manufacturer for repair. At the time of writing we are still awaiting its return and this has introduced some delay in the application of the DSPI system to investigation of real-world problems of practical interest.

7.2 Experimental validation of real-time DSPI instrument

7.2.1 Out-of-plane displacement of a flat test surface

The first prototype of the real-time DSPI system was tested using a flat metal plate target mounted on the Burleigh PZ-81 piezo-electric transducer (PZT) described in section 6.4.2 using double-sided adhesive tape. A second flat metal plate was positioned slightly behind the moving plate to act as a reference surface. A coat of matt white paint was applied to each surface to reduce specular reflections and increase the average reflectivity. A signal generator was used to drive a Burleigh RG93 high voltage amplifier connected to the PZT, producing periodic displacement of the plate along the camera axis. The Mach-Zehnder fibre optic speckle interferometer shown in figure 6-9 was configured in an out-of-plane arrangement and used to acquire phase-stepped interferograms. A 28-80 mm Nikon zoom camera lens was set to a focal length of 50 mm and used to create an image of the two plates on the CAM17 CCD sensor with a field of view of approximately 56 mm x 56 mm. Images were acquired using an aperture ratio of f/8 with an effective magnification factor of approximately 0.14. The image plane speckle size σ can be computed by equation A-6 (see appendix A1) where \( \lambda = 532 \) nm is the illumination wavelength, \( F = 8 \) is the aperture ratio, and \( M = 0.14 \) is the magnification at which the lens is operating, and is approximately equal to 12 \( \mu \)m. The CCD pixel size is 15 \( \mu \)m x 15 \( \mu \)m, resulting in less than two speckles per pixel site.
Application of the real-time digital speckle pattern interferometry instrument computed by the Advanced Processor module located on one of the MAX250 VME boards was sometimes in error. Investigations by the author revealed that one of the arithmetic bits in the result was not being set correctly and this was traced to a poor electrical connection between the main printed circuit board and the plug-in module. Cleaning of the electrical connector and re-seating the module finally resolved the problem. More recently, one of the MV250 boards has developed a hardware fault associated with image memory requiring that the device be shipped to the manufacturer for repair. At the time of writing we are still awaiting its return and this has introduced some delay in the application of the DSPI system to investigation of real-world problems of practical interest.

7.2 Experimental validation of real-time DSPI instrument

7.2.1 Out-of-plane displacement of a flat test surface
The first prototype of the real-time DSPI system was tested using a flat metal plate target mounted on the Burleigh PZ-81 piezo-electric transducer (PZT) described in section 6.4.2 using double-sided adhesive tape. A second flat metal plate was positioned slightly behind the moving plate to act as a reference surface. A coat of matt white paint was applied to each surface to reduce specular reflections and increase the average reflectivity. A signal generator was used to drive a Burleigh RG93 high voltage amplifier connected to the PZT, producing periodic displacement of the plate along the camera axis. The Mach-Zehnder fibre optic speckle interferometer shown in figure 6-9 was configured in an out-of-plane arrangement and used to acquire phase-stepped interferograms. A 28-80 mm Nikon zoom camera lens was set to a focal length of 50 mm and used to create an image of the two plates on the CAM17 CCD sensor with a field of view of approximately 56 mm × 56 mm. Images were acquired using an aperture ratio of f/8 with an effective magnification factor of approximately 0.14. The image plane speckle size $\sigma$ can be computed by equation A-6 (see appendix A1) where $\lambda = 532$ nm is the illumination wavelength, $F = 8$ is the aperture ratio, and $M = 0.14$ is the magnification at which the lens is operating, and is approximately equal to 12 µm. The CCD pixel size is 15 µm × 15 µm, resulting in less than two speckles per pixel site.
Application of the real-time digital speckle pattern interferometry instrument computed by the Advanced Processor module located on one of the MAX250 VME boards was sometimes in error. Investigations by the author revealed that one of the arithmetic bits in the result was not being set correctly and this was traced to a poor electrical connection between the main printed circuit board and the plug-in module. Cleaning of the electrical connector and re-seating the module finally resolved the problem. More recently, one of the MV250 boards has developed a hardware fault associated with image memory requiring that the device be shipped to the manufacturer for repair. At the time of writing we are still awaiting its return and this has introduced some delay in the application of the DSPI system to investigation of real-world problems of practical interest.

7.2 Experimental validation of real-time DSPI instrument

7.2.1 Out-of-plane displacement of a flat test surface

The first prototype of the real-time DSPI system was tested using a flat metal plate target mounted on the Burleigh PZ-81 piezo-electric transducer (PZT) described in section 6.4.2 using double-sided adhesive tape. A second flat metal plate was positioned slightly behind the moving plate to act as a reference surface. A coat of matt white paint was applied to each surface to reduce specular reflections and increase the average reflectivity. A signal generator was used to drive a Burleigh RG93 high voltage amplifier connected to the PZT, producing periodic displacement of the plate along the camera axis. The Mach-Zehnder fibre optic speckle interferometer shown in figure 6-9 was configured in an out-of-plane arrangement and used to acquire phase-stepped interferograms. A 28-80 mm Nikon zoom camera lens was set to a focal length of 50 mm and used to create an image of the two plates on the CAM17 CCD sensor with a field of view of approximately 56 mm x 56 mm. Images were acquired using an aperture ratio of f/8 with an effective magnification factor of approximately 0.14. The image plane speckle size $\sigma$ can be computed by equation A-6 (see appendix A1) where $\lambda = 532$ nm is the illumination wavelength, $F = 8$ is the aperture ratio, and $M = 0.14$ is the magnification at which the lens is operating, and is approximately equal to 12 $\mu$m. The CCD pixel size is 15 $\mu$m x 15 $\mu$m, resulting in less than two speckles per pixel site.
The signal generator was configured to produce a triangular waveform with a period of approximately 3s and the amplitude adjusted to produce a peak-to-peak displacement corresponding to 75% of the full extension range of the PZT (nominally specified at 2 µm). The camera frame rate is fixed at 60 frames s\(^{-1}\) resulting in an inter-frame displacement of approximately 16 nm, equivalent to 0.38 rad of phase error per step. Phase-stepped speckle interferograms were continuously acquired for approximately 5 minutes and processed by the DSPI system in real-time, with the reference phase map updated every 30 s.

Figure 7-1 shows the unwrapped phase difference map \( \Delta \phi(t,0) \) computed in real-time by the prototype DSPI system from 4564 acquired interferograms, corresponding to a total measurement time of approximately 5 minutes. The field of view includes the fixed reference surface to the left of column 130, and the translated surface to the right. It can be seen that the final unwrapped phase difference measured over the reference surface is relatively uniform but includes a considerable offset with respect to the initial state. The mean value computed over columns 1 through 120 is 26.7 rad with standard deviation 0.6 rad. The unwrapped phase values measured across the moving target correspond to the displacement state at the end of the measurement period. The vertical phase gradient observed across the surface of the moving plate on the right hand side suggests some relaxation of the adhesive tape used to bond the aluminium plate to the end surface of the PZT. Small isolated regions of noise are visible in the figure and these are due to internal hardware registers overflowing during the pipeline computation. The prototype system used here performed the spatial speckle averaging as described in the previous chapter, however, the hardware implementation of the neighbourhood multiply and accumulate (NMAC) processing element specified in hardware configurations D1 and D2 only supports operands up to a maximum word length of 8 data bits. Unwrapped phase values are corrupted over a region of 3 x 3 pixels if the NMAC operand extends outside this dynamic range at any time during the measurement period. Data corruption due to register overflow is always a potential problem in systems using fixed-point arithmetic, however, this particular problem has since been resolved by increasing the supported dynamic range of the NMAC operand by a factor of 16. The hardware schematics included in appendix A8 show the latest implementation with an operand word length of 12 bits.
The live unwrapped phase display was recorded on video tape, however, it was not possible to store all of these intermediate images in the system memory due to the large amount of data processed. The limited memory resources were therefore used to store the complete set of intermediate phase values for only a small spatial region of interest that intersected both the reference surface and the moving plate.

Figure 7-2 plots the unwrapped phase change measured during the first 5 seconds of the measurement period for eight pixels. Four adjacent pixels corresponding to the fixed reference plate region are shown, together with four adjacent pixels corresponding to the moving plate region. The pixels are all located on row 250 lying between column 116 and 135, and are indicated by the two short black lines in figure 7-1. The average phase change seen within the fixed plate region is almost constant, whilst within the moving plate region the unwrapped phase follows the triangular waveform of the PZT drive signal. The peak-to-peak phase change is measured as 32.0 rad and the slope of each ramp determined by linear regression is 20.7 rad s\(^{-1}\), corresponding to a signal period of 3.1 s. The measured peak-to-peak surface displacement amplitude is therefore 1.35 \(\mu\)m.

The results from the same experiment are presented over the longer time interval of 5 minutes in figure 7-3 and clearly show some drift in the measured values. Errors introduced by the integer arithmetic pipeline implementation would normally appear as systematic noise and are therefore unlikely to be contributing significantly in the observed result. The random drift is observed in both the fixed plate and moving plate regions and, considering the extended duration of the experiment, is likely to be attributable to air currents occurring between the imaging system and the target. The small variation in the offset between the blue and red traces is probably also due to the relaxation of the adhesive used to mount the moving plate.

These results demonstrate the real-time ability of the prototype system to estimate wrapped phase values from phase-stepped speckle interferograms and perform temporal phase unwrapping. Approximately \(512 \times 512 \times 60 \times 60 \times 5 = 4.5\) Gb of acquired image data was processed during the measurement period, with the live unwrapped phase map displayed on the computer monitor throughout the experiment. The final unwrapped phase map obtained by continuous measurement of out-of-plane
Application of the real-time digital speckle pattern interferometry instrument surface displacement shows good correspondence with the physical parameter, even after a relatively long measurement period (compared to most speckle metrology experiments).

### 7.2.2 Alternative phase-stepping interferometer arrangement

Development of the real-time DSPI system was primarily intended to support the out-of-plane Mach-Zehnder interferometer used in the preceding experiment. Unfortunately, however, the 200 mW Nd:YAG laser manufactured by Lightwave developed a hardware fault resulting in a loss of phase stability. Lightwave engineers were unable to solve the fault and found it necessary to reduce the output power by more than 50% to regain stability. This power loss, coupled with attenuation due to the fibre optic beam delivery system (approximately 50%), resulted in a large reduction in the test surface area that could be sufficiently illuminated and was unacceptably small for practical investigations. At the same time, a second out-of-plane interferometer arrangement developed at Loughborough University and constructed by G.H. Kaufmann became available[97]. This phase-stepping interferometer is based on a Coherent 5 W Nd:YVO₄ Verdi continuous wave laser (532 nm) and uses the Pockels cell phase modulator described in section 6.4.2. The interferometer was dismantled and re-assembled on an optical isolation table for use with the real-time DSPI system. The EEV camera mountings were modified to enable alignment with the new interferometer as shown in figure 7-4. An annotated schematic of the out-of-plane arrangement is shown in figure 7-5. The laser output power can be varied electronically over the range 0.5 W to 5.0 W, enabling the illumination intensity to be easily adjusted to suite a wide range of surface finishes on the target. The beam paths are steered using small mirrors thus avoiding the power loss associated with fibre optic delivery systems.

### 7.2.3 Out-of-plane rotation of a flat test surface

The new out-of-plane interferometer has been used to validate the latest implementation of the real-time DSPI system, which includes the modified NMAC functionality. A slightly enhanced acquisition pipeline now enables a temporal record of intensity data from a small number of pixels in the acquired phase-stepped
Application of the real-time digital speckle pattern interferometry instrument

surface displacement shows good correspondence with the physical parameter, even after a relatively long measurement period (compared to most speckle metrology experiments).

7.2.2 Alternative phase-stepping interferometer arrangement

Development of the real-time DSPI system was primarily intended to support the out-of-plane Mach-Zehnder interferometer used in the preceding experiment. Unfortunately, however, the 200 mW Nd:YAG laser manufactured by Lightwave developed a hardware fault resulting in a loss of phase stability. Lightwave engineers were unable to solve the fault and found it necessary to reduce the output power by more than 50% to regain stability. This power loss, coupled with attenuation due to the fibre optic beam delivery system (approximately 50%), resulted in a large reduction in the test surface area that could be sufficiently illuminated and was unacceptably small for practical investigations. At the same time, a second out-of-plane interferometer arrangement developed at Loughborough University and constructed by G.H. Kaufmann became available[97]. This phase-stepping interferometer is based on a Coherent 5 W Nd:YVO₄ Verdi continuous wave laser (532 nm) and uses the Pockels cell phase modulator described in section 6.4.2. The interferometer was dismantled and re-assembled on an optical isolation table for use with the real-time DSPI system. The EEV camera mountings were modified to enable alignment with the new interferometer as shown in figure 7-4. An annotated schematic of the out-of-plane arrangement is shown in figure 7-5. The laser output power can be varied electronically over the range 0.5 W to 5.0 W, enabling the illumination intensity to be easily adjusted to suite a wide range of surface finishes on the target. The beam paths are steered using small mirrors thus avoiding the power loss associated with fibre optic delivery systems.

7.2.3 Out-of-plane rotation of a flat test surface

The new out-of-plane interferometer has been used to validate the latest implementation of the real-time DSPI system, which includes the modified NMAC functionality. A slightly enhanced acquisition pipeline now enables a temporal record of intensity data from a small number of pixels in the acquired phase-stepped
surface displacement shows good correspondence with the physical parameter, even after a relatively long measurement period (compared to most speckle metrology experiments).

7.2.2 Alternative phase-stepping interferometer arrangement

Development of the real-time DSPI system was primarily intended to support the out-of-plane Mach-Zehnder interferometer used in the preceding experiment. Unfortunately, however, the 200 mW Nd:YAG laser manufactured by Lightwave developed a hardware fault resulting in a loss of phase stability. Lightwave engineers were unable to solve the fault and found it necessary to reduce the output power by more than 50% to regain stability. This power loss, coupled with attenuation due to the fibre optic beam delivery system (approximately 50%), resulted in a large reduction in the test surface area that could be sufficiently illuminated and was unacceptably small for practical investigations. At the same time, a second out-of-plane interferometer arrangement developed at Loughborough University and constructed by G.H. Kaufmann became available[97]. This phase-stepping interferometer is based on a Coherent 5 W Nd:YVO₄ Verdi continuous wave laser (532 nm) and uses the Pockels cell phase modulator described in section 6.4.2. The interferometer was dismantled and re-assembled on an optical isolation table for use with the real-time DSPI system. The EEV camera mountings were modified to enable alignment with the new interferometer as shown in figure 7-4. An annotated schematic of the out-of-plane arrangement is shown in figure 7-5. The laser output power can be varied electronically over the range 0.5 W to 5.0 W, enabling the illumination intensity to be easily adjusted to suite a wide range of surface finishes on the target. The beam paths are steered using small mirrors thus avoiding the power loss associated with fibre optic delivery systems.

7.2.3 Out-of-plane rotation of a flat test surface

The new out-of-plane interferometer has been used to validate the latest implementation of the real-time DSPI system, which includes the modified NMAC functionality. A slightly enhanced acquisition pipeline now enables a temporal record of intensity data from a small number of pixels in the acquired phase-stepped
Application of the real-time digital speckle pattern interferometry instrument interferograms to be stored in a ring-buffer located in the 636 kb of unallocated pipeline processor image memory. This feature enables post-processing of the stored data using alternative phase unwrapping algorithms to validate the real-time results generated by the pipeline processor. For the following experiments, scaling of the 24-bit unwrapped phase values was chosen to represent the range \([-256\pi, 256\pi\) rad. For out-of-plane displacement measurement, this corresponds to a dynamic range of \([-34.0, 34.0\) µm.

The system was used to measure the out-of-plane displacement of a flat aluminium disk of diameter 110 mm rotating at a constant angular velocity about an axis perpendicular to the camera axis. The surface of the aluminium disk was blasted with sand particles to create an optically rough surface, however, no other surface preparation was applied. The disk was mounted on a customised rig (shown in figure 7-6) that enables rotation about a vertical axis passing through the centre of the disk and lying in the plane of the test surface. The disk is spring-loaded against the end surface of a Physik Instruments P820.30 piezo-electric transducer, such that extension of the transducer results in a small rotation of the disk. The PZT produces a nominal extension of 450 nm V\(^{-1}\) and can be position at varying radii from the rotation axis. A PTI TT210 function generator was configured to produce a triangular waveform with a long time period (> 30 s) and was connected to the input of the RG93 high voltage amplifier. The amplifier output was stepped down to the safe working voltage range of the PZT device using a simple potential divider circuit with a measured gain of 0.124.

The Coherent laser output power was set to 0.6 W and the object beam arranged to provide uniform illumination across the disk surface. A Nikon camera lens with a focal length of 50 mm was used to create an image of the disk surface on the CCD sensor, with a field of view of approximately 105 mm × 105 mm (represented by the red rectangle superimposed on the photograph in figure 7-6). Images were acquired using an aperture ratio of f/16 with an effective magnification factor of approximately 0.07. The image plane speckle size \(\sigma\) computed by equation A-6 (see appendix A1) is approximately 22 µm, resulting in roughly half a speckle per pixel.
The acquisition pipeline ring-buffer was set-up to store intensity values measured at all pixels along row $n=250$ from the sequence of acquired phase-stepped interferograms. The limited memory available for the ring-buffer therefore restricts the maximum measurement period to approximately 21 s before data stored in the buffer is overwritten.

The PZT was mounted at a radius of 42 mm from the rotation axis and the function generator adjusted to produce a waveform of amplitude 400 V$_{p-p}$ measured at the RG93 amplifier output. The time constant was increased as far as the generator would allow, producing a measured voltage gradient of 5.3 V s$^{-1}$. This corresponds to a voltage ramp of 650 mV s$^{-1}$ applied to the PZT producing a nominal disk angular velocity of 7.0 µrad s$^{-1}$.

Figures 7-7 and 7-8 show the wrapped and unwrapped phase difference maps, respectively, computed by the real-time DSPI system from 1276 phase-stepped interferograms measured over approximately 21.3 s. The reference phase map was not updated during the phase evaluation. The phase maps are a true representation of the DSPI system results; no post-processing to filter or smooth the data has been applied. The rotation axis is aligned with column $m=255$, with the region to the left moving away, and the region to the right moving toward the camera. The wrapped fringes are vertically aligned across the surface of the flat plate and correspond to the surface displacement contours, however, some distortion of the fringes can be seen in the corners of the phase map and this had been caused by the mounting structure at the disk perimeter.

Intermediate wrapped phase values corresponding to pixels along row 250 have been calculated offline using Matlab based on the intensity data stored in the acquisition ring-buffer. The Matlab wrapped phase estimator uses $3 \times 3$ pixel spatial speckle averaging and the same phase re-referencing scheme as employed by the real-time system. Figure 7-9 shows wrapped phase values plotted against the processing frame number for five pixels located in columns 5, 130, 255, 380, and 505. The plots indicate that the temporal sampling rate is more than adequate to satisfy the Shannon sampling criterion even at the disk perimeter. Surface displacement measured near the rotation axis is negligible as expected.
The intermediate wrapped phase values calculated by Matlab have been unwrapped offline using a temporal phase unwrapping algorithm (denoted TPUA) and a spatial phase unwrapping algorithm (denoted SPUA) for comparison against the real-time Datacube result (denoted DQ). The Matlab TPUA and SPUA implementations maintain full arithmetic precision using standard 64-bit floating-point number representation and do not simulate the arithmetic rounding or truncation of intermediate results occurring in the real-time DQ implementation. The SPUA result has been offset by an integer multiple of $2\pi$ to correct for the phase ambiguity inherent in spatial unwrapping methods. The unwrapped phase difference values computed after 319 processing frames along row $n=250$ by each of the Matlab implementations are compared to the real-time results obtained using the pipeline processor in figure 7-10. The TPUA and DQ plots in figure 7-10 have been offset by 50 rad and 100 rad, respectively, to prevent all plots collapsing onto a single line.

The difference between the phase values computed using the SPUA and TPUA Matlab implementations (shown in figure 7-11) is negligible compared to the phase measurement range. The unwrapped phase difference values computed using the real-time temporal phase unwrapping algorithm (DQ) have been compared to those obtained using the offline Matlab temporal phase unwrapping algorithm (TPUA), and the difference plotted in figure 7-12. The mean error calculated across the disk diameter is $-495 \mu$rad with standard deviation 10.8 mrad.

In theory the displacement of the test surface should vary linearly with distance from the rotation axis, assuming that the disk is rigid and does not deform during the measurement period. A straight-line was fitted in a least-squares sense to the unwrapped phase difference values computed using the Matlab TPUA implementation. The phase gradient estimate was computed as $-0.710 \text{ rad} \text{ pixel}^{-1}$, equivalent to a displacement gradient of $-30.1 \text{ nm} \text{ pixel}^{-1}$, and corresponds to a measured angular velocity of $6.9 \mu\text{rad} \text{ s}^{-1}$. The difference between the measured angular velocity ($6.9 \mu\text{rad} \text{ s}^{-1}$) and the expected angular velocity ($7.0 \mu\text{rad} \text{ s}^{-1}$) is likely to be attributable to the PZT performance tolerance ($<5\%$), small errors in the measurement of the PZT position with respect to the rotation axis, and variations in the function generator time constant. It was later found that the PTI function generator
Application of the real-time digital speckle pattern interferometry instrument was particularly unstable at such low frequencies, producing non-linear voltage ramps with large variations in time constant and this is therefore the most likely source of discrepancy in the angular velocities. Deviation of the real-time result (DQ) from this straight-line is plotted in figure 7-13. Considering the form of the error curve and the fact that the surface is stationary at the axis of rotation it is reasonable to assume that the fitted line has a negative bias, which gives rise to the positive errors around the central region of the disk (i.e. 100 ≤ m ≤ 400). Taking this into account then, the amplitude of the phase change error at the disk perimeter is approximately -2.25 rad, corresponding to 1.3% of the magnitude of final unwrapped phase change (i.e. -38 dB). The form of the error curve and polarity of the deviation from the straight-line is consistent with variation in out-of-plane sensitivity vector across the field of view. In the optical arrangement used here, the object beam was introduced at approximately 15° to the camera axis, a distance of 530 mm from the test surface, resulting in a ±1.3% variation in sensitivity vector between the centre and the outer extents of the field of view.

Out-of-plane tilt of the test surface results in lateral translation of the speckle pattern across the pupil plane in a mirror-like fashion[157]. In the optical arrangement used here the test surface is located at a distance of 690 mm in front of the objective lens, and therefore an angular velocity of 6.9 µrad s⁻¹ produces a lateral translation velocity of approximately 9.5 µm s⁻¹ in the pupil plane. The circular aperture has diameter 3.13 mm (i.e. corresponding to f/16) resulting in a speckle decorrelation factor ρ of approximately 6% in the image-plane between the initial and final states.

The function generator was adjusted to increase the waveform amplitude to 800 V<sub>pp</sub> measured at the RG93 amplifier output, corresponding to a nominal angular velocity of 14 µrad s⁻¹, and the experiment repeated. The wrapped phase difference map computed by the real-time DSPI system from 1276 phase-stepped interferograms measured over approximately 21.3 s is shown in figure 7-14. The reference phase map was not updated during the phase evaluation. Vertical fringes are no longer clearly visible in the wrapped phase map, however, this is unlikely to be attributable to spatial aliasing as the angular velocity has only been increased by a factor of 2 compared to the previous experiment. The effect may be attributable to partial speckle
Application of the real-time digital speckle pattern interferometry instrument

decorrelation in the pupil plane (approximately 12%) over the measurement period. The intermediate wrapped phase values computed offline based on the intensity data stored in the acquisition pipeline ring-buffer is shown in figure 7-15. Again we see that the temporal sampling rate satisfies the Shannon sampling criterion even in regions near to the disk perimeter. The unwrapped phase difference values computed along row $n=250$ using the Matlab TPUA and SPUA implementations are compared to the real-time result in figure 7-16. The TPUA and DQ plots have been offset by 100 rad and 200 rad respectively. It can be seen that the simple Matlab spatial unwrapping algorithm completely fails. The Matlab temporal phase unwrapping algorithm and real-time result continue to show good agreement, however, both exhibit unwrapping errors in regions with high surface velocity.

The experiment was repeated using the same angular velocity, however, the reference phase map was now updated every 50 processing frames (equivalent to 3.3 s). Figure 7-17 shows the wrapped phase difference map computed by the real-time DSPI system from 1276 phase-stepped interferograms measured over approximately 21.3 s. In this case phase noise has been significantly reduced and vertical fringes are almost visible once again. The effect of updating the reference phase map can be seen on the intermediate wrapped phase values in figure 7-18, where the wrapped phase difference returns to zero every 50 processing frames. The unwrapped phase difference values computed in Matlab are once again compared to the real-time result in figure 7-19. The TPUA and DQ plots have been offset by 100 rad and 200 rad, respectively. Noise present in the final wrapped phase difference values continues to prevent successful unwrapping using the spatial algorithm. The temporal unwrapping algorithms again show good agreement, and regular updates to the reference phase map has eliminated unwrapping errors in regions of high surface velocity.

The phase gradient estimates computed from the Matlab TPUA plots shown in figures 7-16 and 7-19 are -1.80 rad pixel$^{-1}$ and -1.71 rad pixel$^{-1}$, respectively. These values correspond to measured angular velocities of 17.4 µrad s$^{-1}$ and 16.6 µrad s$^{-1}$, respectively, with the discrepancy most likely attributable once again to drift in the PTI function generator time constant.
The time period of the triangular waveform was adjusted at the function generator to increase the voltage gradient measured at the RG93 amplifier output to \(40 \text{ V s}^{-1}\), corresponding to a voltage ramp of \(5.0 \text{ V s}^{-1}\) applied to the PZT. The resulting nominal angular velocity was therefore increased by almost 4 times to \(54 \mu\text{rad s}^{-1}\). The experiment was repeated, however, the limited dynamic range of the function generator output combined with the steep voltage gradient required that the measurement duration be reduced to 250 processing frames (i.e. 16.7 s). Speckle decorrelation in the pupil plane between the initial and final states is almost 40%. Results are presented using a re-referencing interval of 50 processing frames. Figure 7-20 shows the intermediate wrapped phase values computed offline, where the wrapped phase difference values measured near the disk perimeter exhibit low frequency modulation, which suggests that the Shannon sampling criterion is no longer satisfied. The unwrapped phase difference values computed using each implementation are shown in figure 7-21, and in this case have not been offset. It can be seen that the Matlab and pipeline processor implementations of the temporal phase unwrapping algorithm breakdown in the same way as the surface velocity increases towards and beyond the sampling limit. In practice, this effectively results in the unwrapped phase difference map being wrapped into the phase range \([-\phi_{\text{lim}}, \phi_{\text{lim}}]\) defined by the sampling frequency \(f_s\). In order to satisfy the Shannon sampling criterion for the temporal phase unwrapping algorithm the phase difference measured between processing frames must lie in the range \([-\pi, \pi]\), therefore defining the theoretical upper limit for the current system as \(15\pi \approx 47 \text{ rad s}^{-1}\). For a measurement period of 16.7 s this corresponds to an unwrapped phase limit of \(\phi_{\text{lim}} = 785 \text{ rad}\). However, it can be seen from figure 7-21 that the temporal phase unwrapping algorithm actually begins to breakdown at just over 400 rad, corresponding to a phase gradient of approximately 24 rad s\(^{-1}\) (nearly half the theoretical limit). This phase error can most likely be attributed to the high velocity of the test surface, which results in significant out-of-plane displacement occurring during acquisition of the four phase-stepped interferograms required for each processing frame. Phase-shifting algorithms that offer improved resistance to errors due to mis-calibration of the phase modulator (e.g. the five-frame algorithm described in section 2.2.2 or the Carré...
Application of the real-time digital speckle pattern interferometry instrument algorithm described in section 2.2.3) could be expected to give better performance close to the Nyquist limit.

7.3 Defect detection in composite materials

7.3.1 Vacuum inspection hood

An inspection hood has been constructed that enables pressure loading of flat test surfaces and small samples (see appendix A11 for technical drawings). The hood consists of a transparent 300 mm × 300 mm Perspex (polymethyl methacrylate) viewing window fastened to the front surface of a square aluminium frame and sealed using a low pressure O-ring. The viewing window is inclined at 5° with respect to the back surface of the frame to prevent the illumination beam reflections from the Perspex sheet from entering the camera aperture. The aluminium frame includes standard pipe fittings to enable a vacuum pump to be connected via plastic hosing. When drawing a vacuum, the frame can be attached to flat test surfaces in a “limpet-like” fashion, with a second O-ring on the back surface ensuring a tight seal (see figure 7-22(a), for example). A solid flat aluminium plate has also been machined that can be attached to the back surface of the frame to create a small vacuum chamber (300 mm × 300 mm × 10 mm) suitable for inspection of small samples (as shown in figure 7-22(b), for example).

7.3.2 Out-of-plane displacement of vacuum loaded carbon fibre sample

The vacuum hood described in section 7.3.1 was used to apply low-pressure vacuum loading to a small carbon fibre sample with dimensions 150 mm × 100 mm × 1.5 mm. The sample was constructed by David Panni (currently working in the Structural Integrity Research Group at Loughborough University), and consists of 6 layers of carbon fibre/epoxy prepreg. The material was supplied by Advanced Composites and consists of Toray 300 carbon fibres prepregnated with an epoxy resin that cures at low temperatures (identified by the proprietary code LTM26/T300 250g 52° o'f). The principle axes of each layer were arranged in the following sequence 0°/90° 90°/0° with respect to the shorter side of the sample. An artificial delamination was produced between the two middle layers by inserting a PTFE shim with dimensions 40 mm × 40 mm. The uncured composite lay-up was placed in a
Application of the real-time digital speckle pattern interferometry instrument

steel jig coated in a silicone release agent and loaded using a mechanical press to consolidate the layers, removing small air bubbles and producing a uniform thickness. The sample was heated to 80 °C for 5 h whilst mounted in the press to complete the consolidation process and cure the epoxy.

The sample was loaded into the inspection hood and supported by a small aluminium cross-beam sitting on two bolts screwed into the aluminium backplate, as shown in figure 7-22(b). The sample was therefore able to deform without the constraints of a fixed clamping arrangement. A thin layer of white emulsion was applied to the surface of the black carbon fibre sample to increase the reflectivity.

The out-of-plane displacement of the test surface was measured using the real-time DSPI system as the pressure within the inspection hood was varied. Experiments were conducted using positive and negative pressure gradients, however, we restrict ourselves here to the negative pressure gradients for brevity. The Coherent laser output power was set to 0.6 W and the object beam arranged to provide uniform illumination across the test surface. A Nikon camera lens with a focal length of 50 mm was used to create an image of the disk surface on the CCD sensor, with a field of view of approximately 117 mm × 117 mm. Images were acquired using an aperture ratio of f/11 with an effective magnification factor of approximately 0.07. The image plane speckle size σ computed by equation A-6 (see appendix A1) is approximately 15 µm, resulting in roughly one speckle per pixel.

Air was pumped out of the inspection hood to create a vacuum of 250 mbar with respect to atmospheric pressure. A tap on the pressure hose was closed and the pump switched off to reduce vibration induced in the inspection hood. The tap was then opened slightly to allow air to leak back slowly into the inspection hood. The real-time DSPI measurement was started when the vacuum reached 200 mbar and was terminated when the vacuum had fallen to 10 mbar. A total of 4112 phase-stepped interferograms were acquired corresponding to a measurement period of approximately 69.2 s. If we assume that the rate of pressure change (dP/dt) is proportional to the vacuum (P) with respect to atmospheric pressure (P₀), then we can write.
Application of the real-time digital speckle pattern interferometry instrument

\[
\frac{dP}{dt} = -k(P - P_a),
\]

(7-32)

where \( k \) is the pressure change coefficient determined by the tap control. The pressure change coefficient can therefore be computed as,

\[
k = \frac{1}{t} \ln \left( \frac{P - P_a}{P - P_i} \right),
\]

(7-33)

where \( P_i \) is the initial vacuum. Substitution of the parameters indicates that the pressure change coefficient used in this experiment was approximately \( 4.3 \times 10^{-2} \, \text{s}^{-1} \).

Figures 7-23 and 7-24 show the wrapped and unwrapped phase difference maps, respectively, computed in real-time by the pipeline processor. Spatial speckle averaging was performed over a 3 \( \times \) 3 pixel region and the reference phase map was updated every 100 processing frames. No post-processing has been applied to the phase data. The sharp linear discontinuities observed in the top, right, and lower regions of the phase maps correspond to the boundaries of the rectangular carbon fibre sample and the aluminium backplate. As in the previous results, positive unwrapped phase values correspond to an out-of-plane displacement away from the camera. A horizontal phase gradient can be seen across the surface of the carbon fibre sample corresponding to a global out-of-plane tilt, superimposed on to which local deformation is clearly visible. The local deformation appears to correspond to two independent defects within the sample corresponding to the extents of the PTFE shim.

One possible interpretation of these results is that there is some adhesion between the carbon fibre sheets and the PTFE shim in the central region that is restricting the local deformation.

A second experiment was conducted in which the vacuum was increased to 550 mbar with respect to atmospheric pressure. Air was allowed to leak back into the inspection hood at a controlled rate as before, measuring out-of-plane displacement over the pressure range 500 mbar to 350 mbar. A total of 4592 phase-stepped interferograms were acquired corresponding to a measurement period of approximately 77.2 s. The pressure change coefficient is therefore approximately \( 4.6 \times 10^{-3} \, \text{s}^{-1} \). The wrapped and unwrapped phase difference maps are shown in figures 7-25 and 7-26, respectively. Local deformation observed on the surface of the carbon fibre sample due to the increased vacuum now corresponds to a single large defect. The increase in pressure
Application of the real-time digital speckle pattern interferometry instrument

difference between the air trapped within the delamination and the air surrounding the sample may now be sufficient to break the adhesion between the shim and the adjacent carbon fibre sheets. Note that spatial unwrapping of the tightly packed fringes covering the extents of the PTFE shim is impossible, as the Shannon spatial sampling criterion has been exceeded, resulting in visible aliasing effects. Temporal unwrapping remains feasible, however, since the rate of displacement has not exceeded the Shannon temporal sampling criterion.

It is interesting to note that subsequent experiments at lower vacuums were able to repeat the results shown in figures 7-23 and 7-24, using both positive and negative pressure gradients. Experiments at higher vacuums, such as the one corresponding to figures 7-25 and 7-26, therefore did not cause irreversible damage to the sample or change the nature of the defect. The artificial delamination can therefore be seen to operate in two deformation modes, depending on the vacuum applied.

7.4 Summary

This chapter describes experiments to validate a real-time speckle interferometry analysis system that is capable of calculating and displaying maps of the displacement field at 15 frames s\(^{-1}\). Data from approximately 250,000 pixels are processed using a combination of temporal phase shifting and temporal phase unwrapping. Unwrapping errors due to low modulation pixels are effectively removed by convolution of the data with a 3 x 3 pixel kernel, which restricts spatial cross-talk to immediately adjacent pixels; each pixel cluster therefore behaves as an independent displacement sensor. In the case of spatial unwrapping, by contrast, each pixel is potentially influenced by every other pixel in the field of view. Increases in noise level in the measured displacement field due to speckle decorrelation are reduced by periodically updating the reference phase map.

Results from experiments with an out-of-plane interferometer on test surfaces undergoing controlled rigid body motion have shown robust operation for periods of up to five minutes. Quantitative comparison of results computed in real-time with those calculated offline using Matlab have shown good agreement, and indicate that errors due to fixed-point arithmetic are negligible compared to other error sources.
Out-of-plane surface displacement due to vacuum loading of a small carbon fibre sample containing an artificial delamination was measured by the real-time DSPI system. Results were presented for two experiments using different loading pressures and successfully demonstrate detection of the defect within the sample.
Application of the real-time digital speckle pattern interferometry instrument

7.5 Figures

Figure 7-1: Unwrapped phase difference map $\Delta \phi(t,0)$ measured after 304 s of continuous phase unwrapping, computed by the first prototype of the DSPI system in real-time. The field of view includes the fixed reference surface to the left of column 130, and the translated surface to the right of column 130. The system estimates wrapped phase values using the standard four-frame phase stepping algorithm and implements the extended temporal phase unwrapping algorithm described in section 6.2 on a pipeline image processing architecture. Spatial speckle averaging is performed over a $3 \times 3$ pixel region, and the reference phase map updated every 30 s. Positive unwrapped phase values correspond to out-of-plane displacement away from the camera. The vertical phase gradient across the moving target is probably due to the relaxation of the adhesive used to bond the metal plate to the end surface of the PZT. Two short black lines have been superimposed on row 250 at either side of the plate boundary to indicate those pixels at which a temporal history of unwrapped phase values was stored.
Figure 7-2: Unwrapped phase difference values, $\Delta \phi(t,0)$, plotted against time for eight pixels located on row 250 of figure 7-1. The red traces relate to four adjacent pixels located in columns 116 through 119 and correspond to the fixed reference plate region. The blue traces relate to four adjacent pixels located in columns 132 through 135 and correspond to the moving plate region.

Figure 7-3: Mean unwrapped phase difference values, $\Delta \phi(t,0)$, plotted against time, calculated using data from the 4 pixels within each region represented in figure 7-2. Random drift of the measured values is most likely attributable to air currents occurring between the imaging system and the test surfaces.
Figure 7-4: Photograph of out-of-plane interferometer based on Coherent 5 W Verdi Nd:YVO₄ continuous wave laser (532 nm) and Leysop Pockels cell phase modulator. The laser beam is split into the object and reference arms of the interferometer in the ratio 90:10, respectively. The beam paths have been superimposed onto the photograph for clarity. The object beam can be seen on the right hand side of the photograph, and is directed onto the test surface using mirrors and a 40x beam expander. The reference beam passes through the phase modulator and spatial filter, before being recombined with the speckle pattern formed by the scattered object beam in the ratio 10:90, respectively. An EEV CAM17 digital camera records the resulting speckle interferograms at 60 frames s⁻¹.
Figure 7-5: Schematic representation of the out-of-plane speckle interferometer arrangement shown in figure 7-4.
The rotation axis lies in the plane of the front surface of the aluminium disk seen in the centre of the image. The disk is spring-loaded against the end surface of a Physik Instruments P810.30 PZT such that extension of the transducer results in a small rotation. The position of the PZT can be adjusted by sliding the rear-mounting rail. The EEV CAM17 camera field of view is represented by the red rectangle superimposed on front surface of the disk. The small die-cast aluminium box contains a high-voltage potential divider circuit used to step down the output from the RG-93 amplifier to a safe operating range.
Figure 7-7: Wrapped phase difference map $\Delta \phi_w(t,0)$ measured after 21.3 s and computed by the DSPI system in real-time. The nominal angular velocity of the disk is 7.0 $\mu$rad s$^{-1}$. The rotation axis lies along column $m=255$ with the region on the left hand side displaced away, and the region on the right hand side displaced toward the camera. Spatial speckle averaging has been performed over a $3 \times 3$ pixel region, and the reference phase map has not been updated during the phase evaluation. No post-processing to filter or smooth the phase data has been performed.
Figure 7-8: Unwrapped phase difference map $\Delta \phi(t,0)$ computed by the DSPI system in real-time. The phase values correspond to the wrapped phase map shown in figure 7-7. Positive unwrapped phase values correspond to out-of-plane displacement away from the camera.
Figure 7-9: Intermediate wrapped phase values plotted against processing frame number for five pixels along row \( n=250 \) of figure 7-8. For clarity, each plot has been offset by an integer multiple of 10 rad and the principle phase range \([-\pi, \pi]\) indicated by dashed lines. The phase values have been computed in Matlab based on intensity data stored in the acquire pipeline ring-buffer. Wrapped phase values computed at pixel \( m=255 \) correspond to the rotation axis and are therefore slowly varying. Pixels located at columns \( m=5 \) and \( m=130 \) correspond to the surface region moving away from the camera. Pixels located at columns \( m=380 \) and \( m=505 \) correspond to the surface region moving toward the camera.
Figure 7-10: Unwrapped phase difference values $\Delta \phi(t,0)$ along row $n=250$ of figure 7-7 computed from 319 processing frames. The unwrapped phase values computed in real-time by the Datacube pipeline processor (DQ) are compared to values obtained by Matlab implementations of a simple spatial phase unwrapping algorithm (SPUA) and temporal phase unwrapping algorithm. For clarity, the TPUA and DQ plots have been offset by 50 rad and 100 rad, respectively.
Figure 7-11: Difference between the SPUA and TPUA plots shown in figure 7-10. The amplitude of the phase errors is negligible compared to the measured phase range.

Figure 7-12: Difference between the DQ and TPUA plots shown in figure 7-10. The mean error measured along row $n=250$ is $-495\,\mu\text{rad}$ with standard deviation $10.8\,\text{mrad}$. These phase errors are small compared to the measured phase range, and are likely to be attributable to arithmetic errors caused by rounding and truncation of intermediate results.
Application of the real-time digital speckle pattern interferometry instrument

Figure 7-13: Deviation of the unwrapped phase values computed by the real-time DSPI system (DQ) from a straight-line fitted in a least squares sense to the Matlab TPUA result shown in figure 7-10.

Figure 7-14: Wrapped phase difference map $\Delta \phi_w(t,0)$ measured after 21.3 s and computed by the DSPI system in real-time. The nominal angular velocity of the disk surface is $14.0 \mu$rad s$^{-1}$. The reference phase map was not updated during the phase evaluation.
Figure 7-15: Intermediate wrapped phase values plotted against processing frame number for five pixels along row \( n=250 \) of figure 7-14. For clarity, each plot has been offset by an integer multiple of 10 rad and the principle phase range \([-\pi, \pi]\) indicated by dashed lines.

Figure 7-16: Unwrapped phase difference values \( \Delta \phi(t,0) \) along row \( n=250 \) of figure 7-14 computed from 319 processing frames. For clarity, the TPUA and DQ plots have been offset by 100 rad and 200 rad, respectively.
The application of the real-time digital speckle pattern interferometry instrument...

Figure 7-17: Wrapped phase difference map \( \Delta \phi_w(t,0) \) measured after 21.3 s and computed by the DSPI system in real-time. The nominal angular velocity of the disk surface is 14.0 \( \mu \text{rad s}^{-1} \). The reference phase map was updated every 50 processing frames (i.e. 3.3 s).

Figure 7-18: Intermediate wrapped phase values plotted against processing frame number for five pixels along row \( n=250 \) of figure 7-17. For clarity, each plot has been offset by an integer multiple of 10 rad and the principle phase range \([-\pi,\pi]\) indicated by dashed lines.
Figure 7-19: Unwrapped phase difference values $\Delta \phi(t,0)$ along row $n=250$ of figure 7-17 computed from 319 processing frames. For clarity, the TPUA and DQ plots have been offset by 100 rad and 200 rad, respectively.

Figure 7-20: Intermediate wrapped phase values plotted against processing frame number for five pixels along row $n=250$. The nominal angular velocity of the disk surface is 53.7 $\mu$rad s$^{-1}$. For clarity, each plot has been offset by an integer multiple of 10 rad and the principle phase range $[-\pi, \pi]$ indicated by dashed lines.
Figure 7-21: Unwrapped phase difference values $\Delta \phi(t,0)$ along row $n=250$ computed from 250 processing frames. The nominal angular velocity of the disk surface is 53.7 $\mu$rad s$^{-1}$. The unwrapped phase difference values computed by the pipeline processor and Matlab implementations of the temporal phase unwrapping algorithm show good agreement, and breakdown in the same way as the surface velocity approaches and exceeds the Shannon sampling criterion.
Figure 7-22: Photographs of the inspection hood used to apply vacuum loading to test surfaces: (a) the inspection hood can be attached to large flat samples in a “limpet-like” fashion, where a O-ring seal on the back surface ensures a tight seal with the test surface. The test surface shown is a damaged carbon fibre panel from the underside of a military aircraft wing; (b) a solid aluminium backplate can be attached for inspection of small test surfaces. The test surface shown is the carbon fibre sample discussed in section 7.3.2, where the front surface has been spray-coated with an emulsion of white titanium dioxide developer powder to increase the reflectivity.
Application of the real-time digital speckle pattern interferometry instrument

Figure 7-23: Wrapped phase difference map $\Delta \phi_W(t,0)$ measured after 69.2 s and computed by the DSPI system in real-time. The phase map corresponds to out-of-plane displacement of the surface of a small carbon fibre sample containing an artificial delamination. Positive unwrapped phase values correspond to out-of-plane displacement away from the camera. Vacuum loading applied to the sample decreased from 200 mbar to 10 mbar over the measurement period. The sharp linear discontinuities observed in the top, right, and lower regions correspond to the boundaries of the rectangular sample and aluminium backplate.
Figure 7-24: Unwrapped phase difference map $\Delta \phi(t, 0)$ computed by the DSPI system in real-time. The phase values correspond to the wrapped phase map shown in figure 7-23.

Figure 7-25: Wrapped phase difference map $\Delta \phi_w(t, 0)$ measured after 77.2 s and computed by the DSPI system in real-time. Vacuum loading applied to the sample decreased from 500 mbar to 350 mbar over the measurement period.
Figure 7-26: Unwrapped phase difference map $\Delta \phi(t,0)$ computed by the DSPI system in real-time. The phase values correspond to the wrapped phase map shown in figure 7-25.
Section V: Discussion
8 Discussion and conclusions

The development of whole-field optical methods in recent years has provided engineers with a toolbox of experimental techniques for measuring the basic physical and mechanical properties of materials that are essential in a manufacturing environment. However, these techniques have seen limited industrial adoption compared with conventional methods (e.g. strain gauging and contact measurements) since interpretation of the resulting wrapped phase data by traditional spatial unwrapping algorithms has required expert knowledge and long computation times when inspecting real-world test specimens.

Temporal phase unwrapping provides an alternative approach that simplifies the data analysis by using a time series of wrapped phase maps to unwrap the data along the time axis. This provides significant advantages for applications in which object discontinuities are present as unwrapping errors due to noise and specimen boundaries do not propagate spatially. However, recording and subsequent analysis of the large data sets required by the temporal phase unwrapping algorithm have restricted its viability to research laboratories with minimal constraints on computing equipment and analysis time.

This thesis has investigated the feasibility of applying real-time temporal phase unwrapping to interferograms acquired at video rates of 60 frames s\textsuperscript{-1} from whole-field optical metrology applications, including profilometry by projected sinusoidal fringes, and digital speckle pattern interferometry. The ability to process whole-field data in real-time is desirable as it enables workers in the field to obtain immediate quantitative feedback from experiments, allowing the experimental techniques to become more intuitive.

A study of the temporal phase unwrapping algorithm implemented on a modern personal computer based on the von Neumann architecture has suggested that attainable data throughput rates are inadequate for a practical real-time system. Investigation of other hardware technologies has revealed that pipeline image processing architectures are particularly suited to analysis of two-dimensional image
Discussion and conclusions

data with deterministic real-time performance, and that temporal phase unwrapping of
phase-shifted interferograms acquired at normal video rates can be achieved.

A real-time temporal phase unwrapping system has been developed based on a
commercial pipeline image processor and adapted for surface profile measurement by
projection of a reversed exponential sequence of sinusoidal fringe densities.
Approximately 500,000 co-ordinates can be measured per second and are displayed as
a two-dimensional pseudo-colour image on the computer monitor. Pixels at which the
fringe visibility is low, or at which the sensor has saturated, are automatically
removed using an overlay mask. Discontinuous surfaces are profiled as easily as
continuous ones since each pixel in the image is processed independently of its
neighbours. Depth measurement accuracy is better than 1 part in 4,000 of the
measurement volume dimension using a maximum fringe density of 32 fringes. Errors
introduced by fixed-point arithmetic on the pipeline processor do not appear to be
significant compared to other error sources.

A systematic experimental study of phase error performance in profilometers based
on the projection of white light fringes formed by a spatial light modulator (SLM) has
been conducted in which the real-time constraint was relaxed. Several issues were
investigated including: the choice of phase-shifting algorithm, the choice of temporal
phase unwrapping algorithm, and the effect of projector defocus. The position of the
projector focal plane relative to the measurement volume was found to be the most
critical of these parameters. The attainable precision represents a compromise
between the systematic errors introduced by the pixel structure of the projector SLM,
and the random errors introduced by electronic noise. A measurement precision (at
the 1 \( \sigma \) level) of 1 part in 20,000 of the measurement volume dimension was
achieved when all of these parameters were optimised. Experiments to study the
repeatability of relative distances measured parallel to the \( z \)-axis using the optimised
system have suggested that a tolerance of 10 \( \mu \)m should be attainable provided that
sufficient care is taken to avoid variations in projector magnification due to thermal
expansion. It has been demonstrated that over-exposure of the photodetector
introduces significant measurement errors, and that these can be reduced by use of a
phase-shifting algorithm resilient to higher harmonics.
Discussion and conclusions

The real-time system has been extended to support temporal phase unwrapping of interferograms acquired from digital speckle pattern interferometry (DSPI) applications. Profilometry and DSPI measurements are controlled using a single software application, and can be made using the same camera to allow direct spatial registration of results from each method. The DSPI system implements an extended form of the temporal phase unwrapping algorithm that is designed to improve the signal to noise ratio in phase estimates from speckle interferograms, and includes support for phase estimation by spatial speckle averaging, periodic updates of the reference phase map, and temporal least-squares fitting to the unwrapped phase. The system is compatible with a variety of phase modulators and enables phase-stepped interferograms to be acquired at 60 frames s\(^{-1}\). Live results are displayed at 15 frames s\(^{-1}\), and can be represented as either greyscale wrapped phase maps or pseudo-colour unwrapped phase maps.

The real-time DSPI system has been demonstrated by measuring out-of-plane displacements of test surfaces undergoing controlled rigid body motion, and has shown robust operation for periods of up to five minutes. Quantitative comparison of results computed in real-time with those calculated offline using Matlab have shown good agreement, and indicate that errors due to fixed-point arithmetic are negligible compared to other error sources.

Phase maps computed by the real-time system corresponding to the out-of-plane surface displacement of a small carbon fibre sample due to vacuum loading have been presented. The results clearly demonstrate detection of an artificial delamination defect within the sample.

8.1 Future Work

Development of a profilometry and DSPI system based on a commercial pipeline image processor has demonstrated that real-time temporal phase unwrapping can be achieved at normal frame rates, and allows live quantitative results from whole-field optical measurements to be displayed. The feasibility study has also identified several issues that would benefit from further work.
The real-time system has been extended to support temporal phase unwrapping of interferograms acquired from digital speckle pattern interferometry (DSPI) applications. Profilometry and DSPI measurements are controlled using a single software application, and can be made using the same camera to allow direct spatial registration of results from each method. The DSPI system implements an extended form of the temporal phase unwrapping algorithm that is designed to improve the signal to noise ratio in phase estimates from speckle interferograms, and includes support for phase estimation by spatial speckle averaging, periodic updates of the reference phase map, and temporal least-squares fitting to the unwrapped phase. The system is compatible with a variety of phase modulators and enables phase-stepped interferograms to be acquired at 60 frames s\(^{-1}\). Live results are displayed at 15 frames s\(^{-1}\), and can be represented as either greyscale wrapped phase maps or pseudo-colour unwrapped phase maps.

The real-time DSPI system has been demonstrated by measuring out-of-plane displacements of test surfaces undergoing controlled rigid body motion, and has shown robust operation for periods of up to five minutes. Quantitative comparison of results computed in real-time with those calculated offline using Matlab have shown good agreement, and indicate that errors due to fixed-point arithmetic are negligible compared to other error sources.

Phase maps computed by the real-time system corresponding to the out-of-plane surface displacement of a small carbon fibre sample due to vacuum loading have been presented. The results clearly demonstrate detection of an artificial delamination defect within the sample.

8.1 Future Work

Development of a profilometry and DSPI system based on a commercial pipeline image processor has demonstrated that real-time temporal phase unwrapping can be achieved at normal frame rates, and allows live quantitative results from whole-field optical measurements to be displayed. The feasibility study has also identified several issues that would benefit from further work.
Discussion and conclusions

In the case of the profilometry application, several problems were noted with the current fringe projection system, which uses a commercial business projector based on a digital mirror device (DMD) spatial light modulator and metal halide bulb. In particular, ageing of the bulb appears to introduce large random fluctuations in the illumination level, which can result in significant phase-errors if a transition occurs during acquisition of the phase-stepped fringe patterns. The variations in intensity could possibly be attributed to ablation of the bulb electrodes, however, further research is required to validate this hypothesis. It would be advantageous to detect, and possibly correct, these random transitions within the projector so that phase-errors introduced by affected image sequences could be prevented. Pre-processing of the acquired interferograms to correct for variations in illumination is also possible\[158\], but clearly adds computational overhead and may not be feasible in a real-time implementation. Alternatively, phase-stepping algorithms that measure more intensity samples than the current four-frame implementation provide improved resilience to phase-errors from intensity variations, however, they also result in an extended measurement period. Several new bulb technologies are now emerging in the commercial projector marketplace that may offer more stable ageing properties.

The current implementation relies on the projector to provide low-level control of the spatial light modulator, however, this introduces some problems for the operator. The projector requires that internal picture enhancement controls are set through buttons on the front panel, and that an internal pixel clock used to sample the analogue video input signal is adjusted to match the sampling rate at the video generator. Inappropriate settings of these controls can results in significant measurement errors. The facility to control the spatial light modulator directly from the computer, without the added complication of picture enhancement operators and external controls would simplify the system management from the viewpoint of the operator, and enable the more advanced features of the DMD to be utilised (e.g. custom temporal switching schemes). Aliasing problems due to mismatched sampling rates at the video generator and the projector could be avoided by the use of an all-digital communication channel between the computer and spatial light modulator, and could take the form of the emerging industry standard digital visual interface (DVI) protocol used by many commercial digital computer monitors.
Chapter 5 identified pixelisation effects caused by the limited spatial resolution of the projector SLM as the critical limiting factor in the measurement resolution that can be attained parallel to the z-axis. It would therefore be very interesting to investigate the measurement performance using a range of spatial light modulators with different native spatial resolutions, and with different display technologies (e.g. poly-silicon liquid crystal display panels, liquid crystal on silicon, and microelectromechanical systems). Such an investigation is particularly interesting from a commercial viewpoint since identification of a cost-effective SLM technology that can increase the measurement precision by even a single order of magnitude would allow direct competition in the mechanical co-ordinate measuring machine (CMM) marketplace. Note, however, that measurement resolution in the x- and y-axes continues to be limited by the spatial resolution of the camera CCD sensor.

The profilometry system computes surface co-ordinates as three two-dimensional parameter maps: \( X(m,n) \), \( Y(m,n) \), and \( Z(m,n) \), where indices \( m \) and \( n \) are used to denote the camera pixel co-ordinates. The current system is only able to compute the depth map, \( Z(m,n) \), in real-time and relies on offline processing to calculate \( X(m,n) \), \( Y(m,n) \), correcting for camera magnification and lens distortion. Although it is clearly desirable to compute the complete co-ordinate set in real-time for some applications, there are many cases where this is not required. However, the availability of pipeline image processing hardware with higher pixel clock rates (e.g. 40 MHz Datacube maxPCI) may enable this feature to be added at reasonable cost in the near future.

The procedure described in chapter 4 for calibrating the profilometry system is relatively time consuming (up to an hour) and only provides an accuracy in the region of 1 part in 1,000 of the measurement volume dimension. This is clearly a limiting factor to the adoption of the measurement system in real-world industrial applications. An alternative procedure described by Schreiber et al.[159] enables self-calibrating measurement (i.e. simultaneous determination of co-ordinates and system parameters) to be achieved. The method is analogous to photogrammetry, where the unknown parameters are over-determined and solved using simultaneous equations. This technique offers the potential for rapid deployment of the system at new measurement
Discussion and conclusions

The current Profilometry system offers immunity to environmental disturbances, and offers a theoretical improvement in the calibration accuracy by a factor of $10^2$.

There are many other areas of study that could be pursued with regard to the Profilometry system, however, these are not regarded as directly relevant to the current project. For example, the current system could be extended to support active control of the projected fringe patterns based on feedback from the acquired images. Such techniques can be used to reduce problems associated with variations in environmental illumination, surface texture and colour, and regions of specular reflection. Post-processing of the measured profile data has not been explored in this investigation, but is extremely important in many industrial applications. Interesting areas of research include global co-ordinate transformation methods, techniques to reliably merge individual surface patches into a single mesh that describes the complete test object, automated dimensioning of test surfaces, and quality control based on surface comparison against a template data set.

The DSPI system has been demonstrated using several experimental set-ups, however, these have all been based on an out-of-plane interferometer arrangement. Application of the system to other arrangements should be relatively straightforward as no changes to the data processing hardware or application software are required. Both in-plane and shearing interferometer arrangements are of particular interest to industry and are used in many practical applications; it would therefore be desirable to validate the system under these circumstances.

All common phase shifting algorithms are sensitive to vibrational disturbances at twice the fundamental frequency, which give rise to intensity distortions both at the fundamental frequency and at three times the fundamental frequency[108]. The current DSPI implementation has been designed around the standard four-frame phase-stepping algorithm due to its computational efficiency, however, as with all phase-stepping algorithms based on steps of $\pi/2$ rad, harmonic components in the intensity signal at three times the fundamental carrier frequency are aliased back on to the fundamental. The current system is therefore susceptible to low-frequency environmental vibration, particularly around the fundamental (15 Hz) and at twice the fundamental (30 Hz), making deployment in industrial environments difficult without
the use of a vibration-isolating optical table. This practical limitation may be overcome by increasing the fundamental frequency\[97\] and implementing alternative phase-stepping algorithms, which although equally sensitive to vibration at twice the fundamental, can provide improved immunity to vibration at other frequencies. The seven-bucket algorithm, for example, is relatively insensitive to low-frequency vibrations below the fundamental frequency.

Temporal least-squares fitting to the unwrapped phase enables improvements in the signal to noise for measurements of physical quantities that follow a known characteristic as the test surface is loaded. The fitting function provided by the current DSPI implementation is somewhat limited in that the incremental unwrapped phase measured between reference phase updates is assumed to increase linearly with time. Support for higher order polynomial models is certainly desirable, as is fitting intervals that are independent of the re-referencing interval. The recursive least-squares estimation approach could be substituted by a Kalman filter, which has the advantage of computing an ongoing measure of its confidence in the estimate. The Kalman filter can be used to estimate the current state given past observations (i.e. filtering the measured data), or improve earlier estimates (i.e. smoothing). The so-called "extended Kalman filter" would be suitable for measurements of non-linear physical systems, and applies a linear model around the estimated state.

The DSPI system displays real-time two-dimensional phase maps at 15 frames s\(^{-1}\). allowing the user to select either a wrapped phase, or unwrapped phase representation. However, for many applications it would be advantageous to convert unwrapped phase data to the physical quantity of interest. Furthermore, the current system could be extended to perform two-dimensional convolution of the unwrapped phase map with an appropriate square filter, allowing strain fields to be displayed in real-time for deformation studies.

In practical measurement situations it is often desirable to make a persistent recording of the experimental results, and this is often achieved by writing the resulting coordinate information or final unwrapped phase map to the computer hard disk drive. In the current system, this is a relatively time consuming operation, requiring up to 0.5 s, which is comparable to the total time required for a surface profile
Discussion and conclusions

measurement. Whilst this may be acceptable for one-off measurements, the delay introduced when recording measurement sequences becomes a problem. In the case of DSPI measurements, it is often desirable to make persistent recordings of the intermediate phase data, however, the delay introduced by writing data to disk is likely to compromise the Shannon sampling requirement of the temporal phase unwrapping algorithm. The current system does enable the operator to use unallocated image memory for persistent storage of intermediate results, however, this is a very limited resource and is insufficient for most practical situations. The approach adopted to solve this problem has been to record the live display information to a conventional video recorder. This enables long sequences of data to be stored with no additional overhead to the processing system. In practice, however, it has been found that there can be a significant loss in picture quality, including considerable colour transformation, and therefore extraction of quantitative data from the video recording is likely to introduce significant errors. Further work to develop an efficient method of recording persistent quantitative results is required, and would probably involve real-time streaming of data directly from the image memories to a dedicated hard disk drive using a custom hardware interface, thus relieving the host CPU of additional overhead.
9 References


48. de Mattia P., Fossati-Bellani V. 'Holographic contouring by displacing the object and the illumination beam', Optics Communications 26 17-21 (1978).


62. Takeda M., Yamamoto H. 'Fourier-transform speckle profilometry: three-
dimensional shape measurements of diffuse objects with large height steps
and/or spatially isolated surfaces', Applied Optics 33(34) 7829-7837 (1994).

63. Kikuta H., Iwata K., Nagata R. 'Distance measurement by the wavelength

64. Barnes T.H., Eiju T., Matsuda K. 'Rough surface profile measurement using
speckle optical frequency domain reflectometry with an external cavity
tunable diode laser', Optik 103(3) 93-100 (1996).

65. Tiziani H.J., Franze B., Haible P. 'Wavelength-shift speckle interferometry for
absolute profilometry using a mode-hop free external cavity diode laser',

66. Kuwamura S., Yamaguchi I. 'Wavelength scanning profilometry for real-time

67. Yamaguchi I., Yamamoto A., Kuwata S. 'Speckle decorrelation in surface
profilometry by wavelength scanning interferometry', Applied Optics 37(28)

68. Kinoshita M., Takeda M., Yago H., Watanabe Y., Kurokawa T. 'Optical
frequency-domain imaging microprofilometry with a frequency-tunable liquid-
crystal Fabry-Perot etalon device', Applied Optics 38(34) 7063-7068 (1999).

69. Takeda M., Aoki T., Miyamoto Y., Tanaka H., Ruowei G., Zhibo Z. 'Absolute
three-dimensional shape measurements using coaxial and coimage plane
optical systems and Fourier fringe analysis for focus detection', Optical

70. Geng Z.J. 'Rainbow three-dimensional camera: new concept of high-speed
References


88. Takeda M., Quan G., Kinoshita M., Takai H., Takahashi Y. 'Frequency-multiplex Fourier-transform profilometry: a single-shot three-dimensional shape measurement of objects with large height discontinuities and/or surface isolations', Applied Optics 36(22) 5347-5354 (1997).


142. Zhao H., Chen W., Tan Y. 'Phase-unwrapping algorithm for the measurement of three-dimensional object shapes', Applied Optics 33(20) 4497-4500 (1994).


Appendices

A1 Speckle effect

Coherent light reflected from an optically rough (or diffusely reflecting) surface generates a three-dimensional interference effect known as the speckle phenomenon. In practice, an optically rough surface can be defined as one for which the standard deviation of the height distribution (usually assumed to be a normal distribution) is much greater than the wavelength of the illumination. In practical optical systems we are usually concerned with the intensity distribution (the so-called speckle field) of the scattered light over a plane, either in free space and formed by an imaging system. There are many good descriptions of speckle properties in the literature, however, this section provides only a brief summary of the important results referred to in this thesis.

The speckle size is taken as the statistical average of the distance between adjacent regions of maximum and minimum brightness in the speckle field, and is determined by the aperture angle subtended at the plane defining the speckle field. The observed speckle size (diameter) in a free-space speckle field (or so-called objective speckle field) formed on a screen at a distance \( L \) by coherent light scattered from a circular region of diameter \( D \) can be approximated by

\[
\sigma_{\text{obj}} \approx \frac{1.22 \lambda L}{D},
\]

where \( \lambda \) is the illumination wavelength. Alternatively, an imaging system (such as a camera, for example) can be used to collect the scattered light and focus it on to a screen forming a so-called subjective speckle field. In this case, the image-plane speckle size can be related to the aperture ratio \( F \) and the magnification factor \( M \) at which the lens is operating.

See for example:


It is often useful to understand the intensity distribution of a speckle field and in this context we must distinguish between fully-developed and partially-developed speckle fields. Interference of light scattered from rough surfaces with height variations greater than the wavelength of the illumination give rise to fully-developed speckle-fields. Conversely, partially-developed speckle fields are formed by the interference of light scattered from surfaces where the height variations is less than the wavelength of light. We concern ourselves here only with fully-developed speckle fields as the statistical properties of partially-developed fields are dependent on the surface height variation.

The intensity of a fully-developed speckle pattern is described by the negative exponential relationship

\[ p(I) = \frac{1}{\langle I_N \rangle} \exp \left( \frac{-I}{\langle I_N \rangle} \right), \quad (A-3) \]

where \( p(I) \, dl \) is the probability that the speckle intensity lies between \( I \) and \( (I + dl) \), and \( \langle I_N \rangle \) is the mean speckle intensity. It can be seen that the most likely sampled speckle intensity is zero and hence there are more dark speckles in the field than speckles of any other brightness.

In speckle interferometry we are concerned with the superposition of a speckle wavefront with either a smooth reference wavefront or a second speckle wavefront, and this gives rise to a modified intensity distribution. We consider first a smooth reference speckle interferometer, in which uniform coherent light from a point source is directed along the axes of the speckle-forming rays. In this case, the probability density function is given by

\[ p(I) = \frac{1}{\langle I_N \rangle} \exp \left( -\frac{I + I'}{\langle I_N \rangle} J_0 \left( 2 \frac{\sqrt{I_I}}{\langle I_N \rangle} \right) \right), \quad (A-4) \]
where \( I_s \) is the intensity of the uniform reference field and \( J_0 \) is the Bessel function of zero order with imaginary argument \( i \). If the average speckle intensity, \( \langle I_s \rangle \), is equal to the reference field intensity, \( I_s \), then equation A-4 simplifies to

\[
p(I) = \frac{2}{I_0} \exp\left( -\left( 1 + \frac{2I}{I_0} \right) J_0\left( 2\sqrt{\frac{2I}{I_0}} \right) \right),
\]

(A-5)

where \( I_0 = 2\langle I_n \rangle = 2I_s \). The intensity distribution described by equation A-5 does not differ significantly from that described by equation A-3, and it can be seen that the most probable speckle intensity is still zero. The most significant difference noticed when a reference field is introduced is that the speckle size increases by a factor of two. This can be understood by considering the simple case without a reference field, in which the speckle size roughly corresponds to the spacing of the interference fringes generated by rays coming from diametrically opposite points at the speckle forming aperture. Conversely, when a strong reference wave is introduced on axis, the principal interference will occur with respect to this central ray, so that the maximum bisecting angle between the interfering rays is halved, thus doubling the interference fringe spacing. The subjective speckle size can therefore be approximated by

\[
\sigma_{sub} \approx 2.44 \lambda F (1 + M).
\]

Speckle interferometry arrangements in which two speckle fields are superimposed, such as those used for in-plane surface displacement measurement for example, do not result in significant changes to the speckle size compared to that of the constituent speckle fields. In the case where two speckle fields of equal mean intensity are combined coherently, the result will be a third speckle pattern with the same speckle size and statistical intensity distribution as the original fields. Alternatively, in the case of two speckle fields of equal mean intensity \( \langle I_n \rangle \) combined incoherently, the intensity distribution is described by

\[
p(I) = \frac{4I}{I_0^2} \exp\left( -\frac{2I}{I_0} \right),
\]

(A-7)

Appendices

where $I_a = 2\left(\frac{1}{\gamma}\right)$. In contrast to the previous result, it can be seen that the probability of dark speckles resulting from incoherent combination of two speckle fields is low.

**A2 Lambertian reflectance**

A Lambertian surface appears equally bright from all viewing directions for a fixed distribution of illumination and a Lambertian surface does not absorb any incident illumination. Lambertian reflectance is also called diffuse reflectance since a Lambertian surface distributes all the incident illumination in all surface directions such that the same amount of energy is seen from any direction. Many matte surfaces can be approximated using Lambertian reflectance.

**A3 Specular reflectance**

A specular surface reflects all incident illumination in a direction that has the same angle with respect to the surface normal but is on the opposite side of the surface normal. 'Mirror-like' surfaces can be approximated using specular reflectance. Glossy surfaces are often modelled using a weighted combination of specular and Lambertian reflectance.

**A4 Number representation**

The number $x$ in binary notation is made up of a set of $B$ bits, given by the vector relationship $x = (x_{B-1}, x_{B-2}, \ldots, x_0)$. This is known as the logical representation, since it refers to the bit patterns within the vector that make up $x$. The arithmetic representation requires a scheme to map the bit patterns into a number upon which arithmetic operations can be performed. This mapping can take one of several forms.

In the fixed-point representation, the $2^B$ values that $B$ bits can represent may all be positive (so called unsigned integer representation). Alternatively, the numbers may be centred on zero, consisting of (nearly) equal numbers of positive and negative numbers. For example, the 2's complement representation can count up to a maximum value of $2^{B-1} - 1$ in the positive direction, and $2^{B-1}$ in the negative direction. Fractional numbers are represented in fixed-point notation by multiplying the value. If this value is a power of two, for example $2^F$, this...
Appendices

Multiplication by a constant is analogous to defining a binary "decimal point" or radix, and allocating the right $F$ bits to represent the fraction. For example, if $B = 8$ and $F = 7$, then the range of values may be represented in the form $x.xxxxxxx$, covering a range of $-1.0000$ (decimal) to $0.9921875$ (decimal). It should be noted that the radix remains fixed within the bit field and thus arithmetic operations on two valid numbers can result in an invalid result that is beyond the range of this fixed-point representation (usually referred to as under- or over-flow).

Floating point representation in effect moves the position of the binary decimal point to change the representation as a function of the number being represented. The $B$-bit field is divided into three subfields: sign ($s$) of width $b_s$ bits, mantissa ($m$) of width $b_m$ bits, and characteristic ($c$) of width $b_c$ bits, where $b_s + b_m + b_c = B$. A number $n$ is represented by $n = (-1)^s m 2^c$. The mantissa is normalised to lie within a certain range, usually $1 \leq m < 2$. The IEEE-754 specification defines the most common implementation of floating point representation, where $B = 32$, $b_m = 23$, $b_s = 1$, and $b_c = 8$.

A5 Manufacturers’ datasheets

The following four pages provide an executive summary of the main VME boards installed in the Datacube pipeline image processing development system. The system comprises of a five slot VME racking unit, which includes a MVME167 single board computer and two Datacube MV250 pipeline image processing boards. The datasheets summarise the functionality of each component and detail the main specifications.
Motorola

MVME167
Single Board Computer

Highlights

Motorola's MVME167 single board computer represents the pinnacle of functionality, flexibility, and performance in a CISC-based system. Based on the most powerful CISC microprocessor available, Motorola's MC68040, the MVME167 combines a microprocessor with the memory management and floating-point units to achieve 26 MIPS at 25 MHz and 40 MIPS at 33 MHz. This outstanding processing speed and floating-point performance makes the MVME167 an ideal solution for scientific and industrial applications.

The MVME167's compatibility with existing M6800 family software, including the UNIX® SYSTEM V/68™ Operating System and the VMexec® Software Development Environment offers CISC-based software environments the ability to realize near RISC performance levels while maintaining object code-compatibility with existing software platforms.
Appendices

Processor

<table>
<thead>
<tr>
<th>Type</th>
<th>MX600C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clock Frequency</td>
<td>26.7 MHz</td>
</tr>
<tr>
<td>MIP (Dimensions L1)</td>
<td>135 x 135 mm</td>
</tr>
<tr>
<td>MIP (Dimensions L2)</td>
<td>75 x 75 mm</td>
</tr>
</tbody>
</table>

Memory

<table>
<thead>
<tr>
<th>Type</th>
<th>Factory Dynamic RAM 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed</td>
<td>25.5 MHz</td>
</tr>
<tr>
<td>ECC</td>
<td>3.5 MHz</td>
</tr>
</tbody>
</table>

Controller

<table>
<thead>
<tr>
<th>Type</th>
<th>VMEbus/Locbus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed</td>
<td>25 MHz</td>
</tr>
<tr>
<td>ECC</td>
<td>3.5 MHz</td>
</tr>
</tbody>
</table>

Capacitor

<table>
<thead>
<tr>
<th>Type</th>
<th>Factory Dynamic RAM 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Values</td>
<td>3.5 MHz</td>
</tr>
</tbody>
</table>

VMEbus (IEEE STD 1014)

- Addressing Capabilities: A16, A14, A12
- Data Transfer Capabilities: D92, D16, D14, D12, B16, B14, UAT
- Adapter: E816P1
- Interface Handler: E817
- Interface Generator: A812h
- System Controller: Yes, supported
- Expansion Mecanism: Yes, supported
- VMEbus Local Bus: Yes

SCSI Bus

<table>
<thead>
<tr>
<th>Controller Type</th>
<th>NC9853/70</th>
</tr>
</thead>
<tbody>
<tr>
<td>Address</td>
<td>5.6 MHz</td>
</tr>
<tr>
<td>Local Bus DMA</td>
<td>Yes, with local bus feature</td>
</tr>
</tbody>
</table>

Ethernet

<table>
<thead>
<tr>
<th>Controller</th>
<th>83509CA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local bus DMA</td>
<td>10.4 Mbit/s</td>
</tr>
</tbody>
</table>

TOD Clock

<table>
<thead>
<tr>
<th>TOD Clock Device</th>
<th>M48278, M8KVRAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ticks</td>
<td>1 tick-2 ticks, 1-sec resolution</td>
</tr>
</tbody>
</table>

In Latin America call:
- Brazil: 55-11-815-4220
- Mexico: 52-281-4942
- In Europe calls:
  - Austria: 43-3-610102-0
  - Belgium: 32-2-716-3382
  - France: 33-1-6674-3560
  - Germany: 49-40-236304-0
  - Italy: 39-2-8202-289
- In the Pacific Area calls:
  - Australia: 61-2-9747-198
  - New Zealand: 61-2-9747-198
- In the United States call:
  - 1-516-714-8000
- In the United Kingdom: 0844-8006-121

Netherlands: 31-30-8762-171

Scandinavia: 46-6-714-8000

Spain: 34-1-339-2461

United Kingdom: 0844-8006-121

People's Republic of China:
- 86-10-641-601
- 86-20-310-9210

Taiwan: 886-2-231-8010

Japan: 81-3-3325-8400

Korea: 82-2-720-0005

Copyright 1994 Motorola, Inc. Phoenix, Arizona. Printed in USA

Motorola In Latin America call:
- Brazil: 55-11-815-4220
- Mexico: 52-281-4942
- In Europe calls:
  - Austria: 43-3-610102-0
  - Belgium: 32-2-716-3382
  - France: 33-1-6674-3560
  - Germany: 49-40-236304-0
  - Italy: 39-2-8202-289
- In the United States call:
  - 1-516-714-8000
- In the United Kingdom: 0844-8006-121

People's Republic of China:
- 86-10-641-601
- 86-20-310-9210

Taiwan: 886-2-231-8010

A7
Complete Image Processing on a Single-slot VME Board

High performance imaging has become a reality, cost-effective, and increasingly widespread technology with a growing number of applications in manufacturing, medicine, R&D, and defense, and other areas. Manipulating images at frame-rates however, can be challenging, handling a huge volume of data requires significant computational horsepower.

Datacube meets this challenge with the MaxVideo 250, a single-slot VME card that can acquire, process, store, and display image data at frame-rates. It derives its power from pipeline processing—a method of specialized parallel processing uniquely suited to the demands of image manipulation. Fast VMEbus transfer is available for applications when access to image data by the host CPU is useful.

Each MaxVideo 250 can be equipped with a module to acquire analog or digital images. It interfaces with a wide variety of sensors including imscam, TOI, multi-tap, or multi-spectral cameras, at a variety of speeds and resolutions.

The MaxVideo 250 uses a collection of specialized computational elements connected sequentially as pixels pass through the pipeline. Operations are performed upon them. Multiple parallel pipelines work synchronously to provide tremendous throughput.

A non-blocking crosspoint switch with 640 MBytes of internal bandwidth allows the computational elements to be re-configured between frames. Linear, non-linear, convolution, histogram and statistical processing operations can be performed in parallel by connecting the processing elements under control of Datacube’s ImageFlow software.

The MaxVideo 250 has six VSIMs (Virtual Surface Image Memories) that provide a total of 7 or 28 MBs of memory. In addition to being surface stores, VSIMs have a crosspoint switch, a statistical processor, an LUT, and an ALU, allowing them to perform specific processing functions at 40 MHz.

The MaxVideo 250’s built-in Analog Generator provides a programmable display and graphics overlay for a variety of resolutions and outputs including grayscale, pseudo-color, and color.

The optional Advanced Processor daughter-card enhances the convolution, statistical analysis, LUT processing and binary morphology capabilities of the MaxVideo 250.

Each MaxVideo 250 can be fitted with an optional MaxModule to enhance the performance of a specific image processing function (e.g. image warping). Even when fully loaded with an acquisition device, Advanced Processor, and a MaxModule, the MaxVideo 250 requires only one VMEbus slot.

The Maxbus ports allow image data interchange with no performance penalty between the MaxVideo 250 and other boards such as Datacube’s X10 Display Controller and MD1 Digital Image Recorder. It also allows cascading of multiple MaxVideo 250s for enhanced processing horsepower.

High Speed Image Access (HSIA) enhances the transfer of frame-rate image data over the VMEbus. In some applications, HSIA allows utilization of the host CPU to augment image processing functionality.

Applications development on the MaxVideo 250 is significantly simplified by Datacube’s ImageFlow. This library of C-callable functions synchronizes multiple data transfers at frame-rates, hiding the complexities of the hardware from the programmer. It is also flexible, with the ability to work with a variety of CPUs and operating systems.

The MaxVideo 250 provides a comprehensive image processing solution—a phenomenal 7 BOPS of processing power, excellent flexibility and versatility—overall price-performance without equal.

- High Speed Image Access provides fast transfer of image data over the VMEbus
- 7 BOPS of processing power
- Cascadable multiple boards for increased processing bandwidth
- Image acquisition, storage, processing, and display in a single VME slot
- 7 or 28 MBs of VSIM memory
- Massive, wide-bandwidth, non-blocking crosspoint switch
MaxVideo 250

Specifications & Features
- Permits convolution, statistical analysis, LUT processing, binary morphology, and many other types of operations
- 16QPS processing power
- 256 MHz pipeline processing
- Single-VME slot form factor, even when fully loaded with Advanced Processor, acquisition modules, and MaxModuls
- Includes 7 MB or 28 MB memory/processors
- High Speed Image Access optimizes VMEbus transfer rate
- Cascadable multiple boards for increased processing bandwidth
- Crosspoint switch provides processor reconfiguration at 640 MB/sec
- Various expansion ports to interface to other MaxVideo devices
- With fewer components and all pins soldered or fastened, the 250 is more rugged than the previous model, with an enhanced MTBF

System Components
- Motherboard
  - Architectural Adapter (AB)
    - VMEbus host interface (HSIA P1/P2 connectors)
    - MILbus ports; P5/P4 and P5/P10 front connections. All 20 MHz
    - Port for optional AS or AD image acquisition module
    - Port for an Advanced Processor (AP) daughtercard
    - Port for one MaxModule daughtercard
  - Configured with 7 MB or 28 MB memory/processors (VSIMs)
  - Crosspoint Switch
    - Wide bandwidth (32X32X8), non-blocking
    - High speed (440 MHz)
    - Reconfigurable by software at frame rates
- Analog Generator (AG) provides graphics and overlay capabilities
  - Pseudocolor, true-color, grayscale
  - 8-bit or without 8-bit graphics overlay at 10 MHz, 20 MHz, 40 MHz
  - Output formats include HSIA YUV, square pixel, 512x512, 1024x1024, 128x128
- Arithmetic Unit (AU) linear, non-linear, and statistical processing
  - 11 bit multipliers
  - Seven 10-bit ALUs
  - Ten 8-bit adders
  - Two statistics processors
  - Two row and column address generators
  - Ten 32-bit data paths; 80 MHz
- Digital Surface Image Memory (VSIM) advanced architecture provides virtual memory storage and 40 MHz processing (6 per board)
  - 64 KB memory/clock per each module
  - Custom GAIN (256 gain)
  - Crosspoint switch, statistical processor, LUT, ALU
  - 28 MB (4x8MB) or 7 MB (1MBx5+2MB) factory configurations

Image Acquisition Modules
- Modules have fully programmable timing: master or slave, horizontal, vertical, composite or pixel clock
- Analog Scanner (AS)
  - Conditions and digitizes analog input signals
  - Digitizes rates all 8-bits from DC to 26 kHz
  - Interfaces to most analog 1D or 2D cameras and sensors
  - Programmable gain, offset, filter selection, and synchronization
- Acquire Digital (AD)
  - Accepts 24-bit input signals
  - Data input 24-bit single end, 12-bit differential
  - Interfaces to most 1D or 2D cameras and sensors
  - Data rates from DC to 20 kHz

Clockwise from top: Motherboard, MaxModule, AP, Acquisition Module

Advanced Processor (AP) enhanced processing resources
- Convolution:
  - 64 point, 8-bit FIR filter
  - 1-D modes: 3x3 kernel, 2x4 kernels, 3x3 Sobel edge extraction
  - 1-D 5x4 filter operation
- Convolutions yield 24-bit normalized result
- Statistical analysis:
  - Histogramming an 8-bit data into 256 (24-bit bins)
  - Feature listing, 512 bins
  - Hough transformations: modified, using a single angle per pass
  - LUT processing: 16x16
- Binary Morphology: 3x3 operations when used with the LUT

MaxModules: one per board (with or without AP device)
- MiniWarper provides nth order warping
- mmNMAC enhances convolution and neighborhood processing
- mmRVP enhances rank value, median, max, min filtering

High Speed Image Access
- Provides sustained VMEbus access of up to 10 MB/sec random or sequential 32-bit LWVOH reads or writes
- VSIMs are of a standard configuration, while the 6th VSIM is an enhanced High Speed Image Access (HSIA) configuration
- HSIA advanced VSIM has two memory banks—one bank can be used for video pipeline transfers, with the other available for high speed image access to the VMEbus
- The banks can be swapped at any time under software control

Environmental
- Operating temperature: 0° to 50° C (32° to 122° F)
- Storage temperature: -40° to 100° C (-40° to 212° F)
- Relative humidity: 10% - 90% (non-condensing)
- Air flow requirement: 50 CFM (minimum)

Physical (fully loaded with acquisition device: AP, MaxModule)
- Height: 0.3" (160 mm)
- Length: 9.19" (233.5 mm)
- Depth: 0.8" (203.3 mm)
- Weight: 24 oz (680 grams)

Additional Information
- For more information about the products mentioned in this document, please refer to the following Datasheet literature:
  - Introducing ImageFlow and ImageFlow Basics
  - MaxVideo 250 Technical Specifications
  - MaxVideo 250 Data Sheet
  - MD1 Family of Digital Image Recorders Datasheet
  - XI Display Controller Datasheet

MaxVideo 250. ImageFlow are trademarks of Dataspace, Inc. Other product names and trademarks are property of respective holders.

(All specifications subject to change without notice) 4/95, LIT-900
Appendices

A6 Schematics for digital camera RS422 line driver module

The following seven pages include detailed diagrams of the digital camera RS422 line driver module designed to support the EEV CAM17 camera. The first two diagrams show a schematic representation of the digital circuit, and are followed by the printed circuit board (PCB) layouts. The width of the double-sided PCBs have been chosen such they do not exceed the camera dimensions, and can be stacked vertically to reduce the required footprint area. Finally, folding and construction details of the metal enclosure are included.
PCB artwork legend: Blue = Component side; Red = Track side; Green = Silk screen.
PCB artwork legend: Blue = Component side; Red = Track side; Green = Silk screen.
Exploded view of camera driver enclosure panels
A7 Pipeline schematics for real-time profilometry system

The following diagrams provide a high-level simplified description of the real-time profilometry system implemented on the pipeline image processor. The processing functions are performed by applying a sequence of hardware configurations denoted S1 through S7 as described in chapter 4. The following diagrams provide schematic representations of each of these hardware configurations.

Legend
Pipeline image processor configuration S1
Pipeline image processor configuration S2

Pipeline image processor configuration S3
Pipeline image processor configuration S4

Pipeline image processor configuration S5
Appendices

A8 Pipeline schematics for real-time DSPI system

The following diagrams provide a high-level simplified description of the real-time digital speckle pattern interferometry system implemented on the pipeline image processor. The processing functions are performed by applying a sequence of hardware configurations denoted D1 through D6 as described in chapter 6. The following diagrams provide schematic representations of each of these hardware configurations.

![Diagram of pipeline image processor configuration (D1)](image)

Legend:

- **MEM**: Image surface store
- **BUS width**: Specified as number of bits
- **Dyadic arithmetic processing elements**: 
- **Arithmetic shift processing element**
- **Arithmetic rounding processing element**: 
- **Arithmetic truncation processing element**
- **Multiplex data paths**
- **Merge data paths**
- **Split data paths**
- **Data lookup table (LUT)**
- **Dyadic logical processing elements**
- **N(r, l, m)**
- **D(r, l, m)**
- **F(r, l, m)**
- **Cam in**: 101(l, m)
- **Disp in**: 16
- **DAC_LUT1**: 24
- **DAC1**

Pipeline image processor configuration (D1)
Pipeline image processor configuration  D2
Appendices

Pipeline image processor configuration D3
Pipeline image processor configuration D4
Pipeline image processor configuration D5
Pipeline image processor configuration D6
Appendices

A9 Microbench™ arrangements used in 200 mW out-of-plane interferometer

The following two pages provide scale drawings of the Spindler and Hoyer Microbench™ arrangements used in the out-of-plane interferometer described in chapter 6. The interferometer is based on a Lightwave 200 mW Nd:YAG (532 nm) continuous wave laser and uses single-mode, polarisation-preserving optical fibres to deliver the reference and object beams. The complete arrangement is mounted on a 300 mm × 300 mm aluminium breadboard together with the laser and camera. This set-up allows the complete optical head to be moved to several measurement positions around the object under test with minimal optical realignment.
Appendices

Note: The mount for the object arm fibre optic receiver is located vertically above the beam-splitter cube but has been omitted from this diagram for clarity.
A10 **Microbench™ arrangement used to analyse dynamics of Burleigh PZ81**

The following scale drawing describes the Microbench™ layout used to form smooth wavefront fringes on the high-speed Kodak Ektapro camera CCD sensor as part of an investigation into the performance of the Burleigh PZ81 piezo-electric transducer (PZT). The three-axis optical mount was adjusted to produce vertical fringes across the CCD, whilst the PZT mounted mirror was used for phase stepping the fringe pattern.
Appendices

A11 Vacuum inspection hood

The following three scale drawings detail the construction of the vacuum inspection hood described in chapter 7. The design is based on a square aluminium frame with the top surface inclined at 5° to the horizontal. Grooves are machined into the top and bottom surfaces to support low-pressure O-ring seals, and tapped bore-holes have been included on the top surface to mount the viewing window.
12 of M6 Tapped Bore Hole
To Depth 6mm

Groove for O-ring on top surface
Groove for O-ring on bottom surface

A36

LOUGHBOROUGH UNIVERSITY
Inspection Vessel

Russell Coggrave

26/6/99

SCALE

A3

Sheet 2 of 3
A12 Publication list

A12.1 Refereed journal papers


A12.2 Papers in conference proceedings


