A self-calibrating phase-shifting algorithm based on the natural demodulation of two-dimensional fringe patterns

Kieran G. Larkin

Department of Physical Optics, School of Physics,The University of Sydney, NSW 2006, Australia, and
Canon Information Systems Research Australia, 1 Thomas Holt Drive, North Ryde, NSW 2113, Australia.

http://www.physics.usyd.edu.au/~larkin/

Abstract: A new method of estimating the phase-shift between interferograms is introduced. The method is based on a recently introduced two-dimensional Fourier-Hilbert demodulation technique. Three or more interferogram frames in an arbitrary sequence are required. The first stage of the algorithm calculates frame differences to remove the fringe pattern offset; allowing increased fringe modulation. The second stage is spatial demodulation to estimate the analytic image for each frame difference. The third stage robustly estimates the inter-frame phase-shifts and then uses the generalised phase-shifting algorithm of Lai and Yatagai to extract the offset, the modulation and the phase exactly. Initial simulations of the method indicate that high accuracy phase estimates are obtainable even in the presence of closed or discontinuous fringe patterns.
©2001 Optical Society of America

References, notes and links

1. P. Carre, "Installation et utilisation du comparateur photoelectrique et interferentiel du Bureau International des Poids et Mesures.," Metrologia **2**, 13-23 (1966).

2. W. R. C. Rowley, and J. Hamon, "Quelques mesures de dyssymetrie de profils spectraux," Revue d'Optique **9**, 519-531 (1963).

3. Y.-Y. Cheng, and J. C. Wyant, "Phase-shifter calibration in phase-shifting interferometry," App. Opt. **24**, 3049-3052 (1985).

4. P. Hariharan, B. F. Oreb, and T. Eiju, "Digital phase-shifting interferometer: a simple error-compensating phase calculation algorithm," App. Opt. **26**, 2504-2506 (1987).

5. Y. Ishii, and R. Onodera, "Phase-extraction algorithm in laser diode phase-shifting interferometry," Opt. Lett. **20**, 1883-1885 (1995).

6. K. G. Larkin, and B. F. Oreb, "Design and assessment of Symmetrical Phase-Shifting Algorithms," J. Opt. Soc. Am. A **9**, 1740-1748 (1992).

7. K. G. Larkin, and B. F. Oreb, "A new seven sample phase-shifting algorithm," SPIE International Symposium on Optical Applied Science and Engineering, San Diego, Gordon M. Brown, Osuk Y. Kwon, Malgorzata Kujawinska, Graeme T. Reid, eds., SPIE Proc. **1755**, California, (1992).

8. Y. Surrel, "Phase stepping: a new self-calibrating algorithm.," App. Opt. **32**, 3598-3600 (1993).

9. P. L. Wizinowich, "Phase-shifting interferometry in the presence of vibration: a new algorithm and system," App. Opt. **29**, 3271-3279 (1990).

10   J. Schwider, O. Falkenstorfer, H. Schreiber, A. Zoller, et al., "New compensating four-phase algorithm for phase-shift interferometry," Opt. Eng. **32**, 1883-1885 (1993).

11   K. Hibino, B. F. Oreb, D. I. Farrant, and K. G. Larkin, "Phase-shifting interferometry for non-sinusoidal waveforms with phase-shift errors," J. Opt. Soc. Am. A **12**, 761-768 (1995).

12   K. Hibino, B. F. Oreb, D. I. Farrant, and K. G. Larkin, "Phase-shifting algorithms for nonlinear and spatially nonuniform phase shifts," J. Opt. Soc. Am. A **14**, 918-930 (1997).

13   L. Mertz, "Real-time fringe pattern analysis," App. Opt. **22**, 1535-1539 (1983).

14   W. W. Macy Jr, "Two-dimensional fringe-pattern analysis," App. Opt. **22**, 3898-3901 (1983).

15   K. H. Womack, "Interferometric phase measurement using spatial synchronous detection," Opt. Eng. **23**, 391-395 (1984).

16   P. L. Ransom, and J. V. Kokal, "Interferogram analysis by a modified sinusoid fitting technique," App. Opt. **25**, 4199-4204 (1986).

17   M. Takeda, H. Ina, and S. Kobayashi, "Fourier-transform method of fringe-pattern analysis for computer-based topography and interferometry," J. Opt. Soc. Am. **72**, 156-160 (1982).

18   M. Takeda, and K. Mutoh, "Fourier transform profilometry for the automatic measurement of 3-D object shapes," App. Opt. **22**, 3977-3982 (1983).

19   K. H. Womack, "Frequency domain desciption of interferogram analysis," Opt. Eng. **23**, 396-400 (1984).

20   K. A. Nugent, "Interferogram analysis using an accurate fully automatic algorithm," App. Opt. **24**, 3101-3105 (1985).

21   D. J. Bone, H.-A. Bachor, and R. J. Sandeman, "Fringe-pattern analysis using a 2-D Fourier transform," App. Opt. **25**, 1653-1660 (1986).

22   C. Roddier, and F. Roddier, "Interferogram analysis using Fourier transform techniques," App. Opt. **26**, 1668-1673 (1987).

23   M. Takeda, and M. Kitoh, "Spatiotemporal multiplex heterodyne interferometry," J. Opt. Soc. Am. A **9**, 1607-1614 (1992).

24   G. Lai, and T. Yatagai, "Generalized phase-shifting interferometry," J. Opt. Soc. Am. A **8**, 822-827 (1991).

25   C. J. Morgan, "Least squares estimation in phase-measurement interferometry," Opt. Lett. **7**, 368-370 (1982).

26   J. E. Grievenkamp, "Generalised data reduction for heterodyne interferometry," Opt. Eng. **23**, 350-352 (1984).

27   T. Kreis, *Holographic interferometry. Principles and methods*, 1, Akademie Verlag GmbH, Berlin, 1996.

28   K. Okada, A. Sato, and J. Tsujiuchi, "Simultaneous calculation of phase distribution and scanning phase in phase shifting interferometry," Opt. Comm. **84**, 118-124 (1991).

29   G. D. Lassahn, J. K. Lassaahn, P. L. Taylor, and V. A. Deason, "Multiphase fringe analysis with unkown phase shifts," Opt. Eng. **33**, 2039-2044 (1994).

30   G.-S. Han, and S.-W. Kim, "Numerical correction of reference phases in phase-shifting interferometry by iterative least-squares fitting," App. Opt. **33**, 7321-7325 (1994).

31   I.-B. Kong, and S.-W. Kim, "General algorithm of phase-shifting interferometry by iterative least-squares fitting," Opt. Eng. **34**, 183-188 (1995).

32   K. E. Perry_Jr, and J. McKelvie, "Reference phase shift determination in phase shifting interferometry," Opt. Lasers Eng. **22**, 77-90 (1995).

33   Tests by the author have shown that the error term to be minimised can vanish for the common four sample algorithm (with nominal phase steps of 90 degrees), even when phase step errors are present. This can be seen as stagnation in the optimisation procedure.

34   E. W. Rogala, and H. H. Barrett, "Phase-Shifting Interferometry and Maximum-Likelihood Estimation Theory - II - a Generalized Solution," App. Opt. **37**, 7253-7258 (1998).

35   C. T. Farrell, and M. A. Player, "Phase step measurement and variable step algorithms in phase-shifting interferometry," Meas. Sci. Techol. **3**, 953-958 (1992).

36   C. T. Farrell, and M. A. Player, "Phase-step insensitive algorithms for phase-shifting interferometry," Meas. Sci. Techol. **5**, 648-652 (1994).

37   A. W. Fitzgibbon, M. Pilu, and R. B. Fisher, "Direct least squares fitting of ellipses," IEEE Transactions on Pattern Analysis and Machine Intelligence **21**, 476-480 (1999).

38   H. van Brug, "Phase-step calibration for phase-stepped interferometry," App. Opt. **38**, 3549-3555 (1999).

39   X. Chen, M. Gramaglia, and J. A. Yeazell, "Phase-shifting interferometry with uncalibrated phase shifts," App. Opt. **39**, 585-591 (2000).

40   J. Schwider, T. Dresel, and B. Manzke, "Some considerations of reduction of reference phase error in phase-stepping interferometry," App. Opt. **38**, 655-659 (1999).

41   K. G. Larkin, and B. F. Oreb, "Propagation of errors in different phase-shifting algorithms: a special property of the arctangent function," SPIE International Symposium on Optical Applied Science and Engineering, San Diego, California, Gordon M. Brown, Osuk Y. Kwon, Malgorzata Kujawinska, Graeme T. Reid , eds., Proc. SPIE **1755**, (1992), 219-227.

42   J. Li, X.-Y. Su, and L.-R. Guo, "Improved Fourier transform profilometry for the automatic measurement of three-dimensional object shapes," Opt. Eng. **29**, 1439-1444 (1990).

43   R. Windecker, and H. J. Tiziani, "Semispatial, robust, and accurate phase evaluation algorithm," App. Opt. **34**, 7321-7326 (1995).

44   K. A. Goldberg, and J. Bokor, "Fourier-transform method of phase-shift determination," App. Opt. **40**, 2886-2894 (2001).

45   K. G. Larkin, D. Bone, and M. A. Oldfield, "Natural demodulation of two-dimensional fringe patterns: I. General background to the spiral phase quadrature transform.," J. Opt. Soc. Am. A **18**, 1862-1870 (2001).

46   G. H. Granlund, and H. Knutsson, *Signal processing for computer vision*, Kluwer, Dordrecht, Netherlands, 1995.

47   K. G. Larkin, "Natural demodulation of two-dimensional fringe patterns: II. Stationary phase analysis of the spiral phase quadrature transform.," J. Opt. Soc. Am. A **18**, 1871-1881 (2001).

48   K. G. Larkin, "Topics in Multi-dimensional Signal Demodulation", PhD. University of Sydney, 2001.

49   T. Kreis, "Digital holographic interference-phase measurement using the Fourier transform method," J. Opt. Soc. Am. A **3**, 847-855 (1986).

50   Explicit orientation estimation can be eliminated from the technique at the expense of clarity. The orientation estimate is replaced by a random but consistent choice of polarity $h(x,y)$. The two levels of differencing in the algorithm ultimately remove the function $h(x,y)$ from the final result.

51   D. Gabor, "Theory of communications," Journal of the Institution of Electrical Engineers, **93**, 429-457 (1947).

52   J. P. Havlicek, J. W. Havlicek, and A. C. Bovik, "The Analytic Image,," IEEE International Conference on Image Processing, Santa Barbara,California, (1997), 446-449.

53   T. Bülow, and G. Sommer, "A Novel Approach to the 2D Analytic Signal," Computer Analysis of Images and Patterns, CAIP'99, Ljubljana, Slovenia, (1999), 25-32.

54   M. A. Fiddy, "The role of analyticity in image recovery", *Image recovery: theory and application*, ed. Stark, H. (Orlando, Florida: Academic Press, 1987).

55   M. Alonso, and G. W. Forbes, "Measures of spread for periodic distributions and the associated uncertainty relations," Am. J. Phys. **69**, 340-347 (2000).

56   C. J. R. Sheppard, and Z. S. Hegedus, "Axial behavior of pupil plane filters," J. Opt. Soc. Am. A **5**, 643-664 (1988).

57   B. Strobel, "Processing of Interferometric Phase Maps As Complex-Valued Phasor Images," App. Opt. **35**, 2192-2198 (1996).

58   B. Zhao, "A statistical method for fringe intensity-correlated error in phase-shifting measurement: the effect of quantization error on the N-bucket algorithm," Meas. Sci. technol. **8**, 147-153 (1997).

59   C. Brophy, "Effect of intensity error correlation on the computed phase of phase-shifting interferometry," J. Opt. Soc. Am. A **7**, 537-541 (1990).

1. Introduction

Automated interferogram analysis can be split into two broad categories: temporal (phase-shifting) methods, and spatial methods. Interestingly, the first phase-shifting algorithms (or PSAs) proposed in the 1960's by Carré [1], and Rowley and Hamon [2] are self-calibrating. One of the enduring problems with temporal PSAs has been the phase-shift calibration (or shift error) problem. Most techniques explicitly rely upon the shift between measurements being an exact, predictable amount. Numerous self-correcting algorithms have been developed over the years [1, 3-10] to compensate for shift errors, originating primarily from the imperfect piezo transducers used to shift the reference mirror. By using considerably more than four frames it is possible for algorithms to compensate deterministic shift errors (such as first, second and third order piezo nonlinearities) and non-sinusoidal fringe profiles [11, 12].

Spatial fringe analysis methods have developed in parallel with the temporal PSAs. The spatial methods were often developed for situations where the multiple phase measurements using temporal PSAs are impossible or impractical. Many spatial methods can be seen as fitting a sinusoid to a fringe profile so that the phase can be estimated at all points – not just the fringe maximum and minimum [13-16]. The fitting of (1-D) phase modulated sinusoids has an elegant Fourier interpretation, first noted by Takeda [17, 18], then several others [14, 19, 20]. In 1986 the analysis was formally extended to 2-D by Bone[21] then Roddier [22]. The method is usually referred to as the Fourier transform method , or FTM for short.

About ten years ago researchers began considering whether or not the spatial and temporal methods could be combined to relax the restriction on the phase-shift between frames. Subsequently a number of sophisticated self-calibrating techniques have been developed. The purpose of this paper is to present a new self-calibrating algorithm which has the potential for very high accuracy calibration using as few as three interferograms. It is also an objective of this work to place the results in the context of the main alternative techniques; briefly reviewed in the following section. However, it has not been possible to make general quantitative comparisons as most other techniques cannot calibrate closed fringe patterns using just three frames.

In section 3 the closely related but almost forgotten FTM calibration method and its limitations are considered in more detail. Section 4 introduces a new isotropic Fourier demodulation technique (called the vortex transform for brevity) which overcomes the problems of the older FTM. In section 5 the vortex transform is applied to phase-shift calibration. Section 6 presents a summary and flowcharts of the full calibration procedure. Some accuracy estimates are derived in section 7. The final section contains a discussion of the results and further applications.

2. Self calibrating spatio-temporal algorithms

It transpires, for typical interferograms, that there is often significant redundancy in the information contained within a series of phase shifted fringe image frames. The redundancy can be taken advantage of in self-calibrating schemes. Another way of viewing this is as a strong localisation in the spatio-temporal frequency domain [23].

2.1 Fourier method

Lai and Yatagai [24] were, perhaps, the first to suggest using a spatial technique to estimate the phase in each frame separately, and then to use a generalized PSA to calculate the phase from all the frames. The generalized algorithm [25, 26] works with 3 or more frames having arbitrary, but known, phase shifts (except for some degenerate combinations that are

statistically rare). The spatial phase method proposed relies on each interferogram frame containing an exactly linear fringe region (viz. a phase wedge) which can be demodulated with the FTM. In reality the method can work with fringes that are not straight and equally spaced, but this idea has only been hinted at by Kreis [27] (on page 146) as a method for removing the sign ambiguity in the FTM. The limitations of this method are discussed further in section 3.

2.2 Overdetermined system of linear equations

An alternative approach using the statistical properties of interferogram sequences was proposed by Okada et al [28] in 1991. The basic idea being that if the number of image pixels and the number of frames are both large enough, then there will be a system of simultaneous equations with more equations than unknowns. In Okada's scheme the equations are solved iteratively after finding an approximate (linear) solution. Consider the nth interferogram frame in a sequence

$$f_n(x,y) = a(x,y) + b(x,y)\cos\left[\psi(x,y) + \delta_n\right].$$
(1)

The function $a(x,y)$ is the background or offset. The function $b(x,y)$ is known as the (amplitude) modulation, and the function $\psi(x,y)$ is the phase. Each interferogram has a phase shift parameter δ_n. It is possible to calculate the value of $\psi(x,y)$ from three or more frames if the value of δ_n is known for each frame. For three frames the value of $\psi(x,y)$ is exactly determined. For more than three frames $\psi(x,y)$ is over-determined. When the values of δ_n are not known, the situation changes. Consider a system consisting of N frames, each containing M pixels. Each pixel in each frame represents an intensity measurement, so there are NM simultaneous equations. The parameters $a(x,y)$, $b(x,y)$, and $\psi(x,y)$ are different (in general) for each of the M pixel locations ($3M$ variables). The shift parameters δ_n only vary between the N frames (or N-1 variables, assuming $\delta_0 = 0$ is taken as reference). Thus the overall number of unknowns is $3M + N - 1$. The usual argument continues: there must be at least as many equations as unknowns to solve this system, hence

$$NM \geq 3M + N - 1 \ \text{ or } \ N \geq \frac{3M - 1}{M - 1} \ \text{ or } \ M \geq \frac{N - 1}{N - 3}.$$
(2)

At this point it is clear that an unstated assumption is that all NM equations are independent. Note that the equality corresponds to an exact solution, whereas the greater than sign corresponds to a least squares type solution. The conclusion is that $N \geq 4$ for typical interferograms (M is large). Lassahn et al [29] proposed a similar method in 1994. Soon after that Han and Kim [30] proposed an iterative least squares fitting technique, followed by Kong and Kim [31]. Perry and McKelvie [32] compared a number of least–squares estimation algorithms, concluding that the initial phase-shift estimates must be within 5° for the final phase to be correct within $\lambda/100$ (3.6°). Although all these methods claim that four frames should be sufficient, no published simulations have used less than five frames ($N \geq 5$) as far as I am aware, so there is still some doubt. It is suspected that there are certain conditions that cause the iterative approach to stagnate at locations far from the ideal solution, although this conjecture has yet to be proved [33].

A recent approach to the problem by Rogala and Barrett [34] seems to have evolved in complete isolation from the least squares fitting methods above. Their derive a maximum likelihood estimate through purely statistical methods and show that several well known PSAs are truly optimal.

2.3 Lissajou ellipse fitting method

The approach of Farrell and Player [35, 36] is a geometric interpretation of Eq. (1). At any two fixed pixel locations Eq. (1) describes an ellipse if one pixel intensity is plotted versus the other intensity (Lissajou figure). Five separate points are required to uniquely define an ellipse, and this corresponds precisely with the previous method: $M = 2 \Rightarrow N \geq 5$. In practice the method requires seven or more frames for reliable ellipse fitting using the Bookstein method. Recently a more robust and efficient method of ellipse fitting has been derived [37] and may be expected to make this approach to phase calibration more attractive.

2.4 Interferogram correlation method

Van Brug [38] has proposed a method based upon the correlation of pairs of interferograms. The idea of using correlation to estimate the phase-shift between interferograms (expressed in Eq. (1)) is closely related to the Fourier transform approach implied by Kreis [27], although van Brug seems unaware of the connection. In the Fourier domain a cross-correlation is simply a multiplication with one function complex conjugated – hence Fourier phases subtract. The main source of errors in this technique is the primary assumption that the interferograms consist of a large number of linear fringes. Any deviation from this assumption invalidates the analysis and immediately introduces errors. Although van Brug goes to great lengths plotting the results of simulations varying the number of linear fringes in a circular interferogram window, it is easy to show that the error is simply related to the Fourier transform of the windowing function (a $J_1(x)/x$ Airy pattern variation). The error can be interpreted as the correlation of two shifted fringe patterns, each of which has an unshifted window. The correlation has two conflicting components: one shifted, one fixed. The algorithm, unlike most others considered, works for as few as two fringe pattern frames. Unfortunately the method only gives tolerably small error for highly linear fringe patterns and so is not useful more generally.

Fringe pattern min-max comparison

Recently Chen et al [39] have proposed a calibration method based on two-point correlations. In essence, two pixels are chosen arbitrarily and the sum and difference of their intensities on each frame is calculated. The sums and differences are calculated for each frame and the overall maximum and minimum are found. It can be shown that the min-max difference is related to the cosine of the phase difference between the points if the fringe patterns intensities have been normalised. The method requires 15 or more well distributed frames to work adequately. So although the algorithm is not appropriate for a small number of frames, it is relatively efficient because only a small number of pixel pairs are needed for calibration. The method can be considered as a special case of interferogram correlation (using only extrema).

2.6 Other methods

Recently Schwider et al [40] have proposed detection of the second harmonic phase errors (which are a characteristic of phase-shift miscalibration [41]). The method is fine for simple, smooth phase patterns (such as linear and quadratic surfaces), but cannot necessarily discriminate errors from phase patterns which have some inherent second harmonic structure. An alternative method has also been proposed for estimating the effect of second harmonics errors on the histogram of the phase map. However this scheme is even more restrictive on the underlying phase pattern in that it requires a uniform histogram. The implication is that the fringes must be substantially linear and equally spaced, and therefore not very interesting.

3. Can all of the spatial information be used to calibrate the temporal phase-shift?

Several authors have suggested using (exact) temporal information to improve the accuracy or range of the spatial phase measuring methods. Li et al [42] proposed the use of two shifted interferograms (with a π radian shift) in the FTM to remove the DC offset and allow wider Fourier sidebands, and hence larger spatial-phase gradients. Windecker and Tiziani [43] combine spatial and temporal methods by using the exact temporal shifts to improve the accuracy of the spatial estimators. Can the situation be turned around so that as all of the spatial information is used to estimate the relative phase-shift of each frame? Until now the answer has been no, in general. A partial solution was mentioned in section 2.1; Lai and Yatagai used spatial information from special linear phase wedges to calibrate the relative phase of each frame. But their method seems too conservative. Why not just use the FTM on each full frame and then calculate the frame differences? This approach seems to be implied by Kreis (in his book [27] on page 146), but no subsequent published work seems to have developed the idea. The method actually works for slightly curved and unequally spaced fringes, so long as the Fourier sidelobes remain distinct and well separated. In fact the method has been recently rediscovered and developed by Goldberg and Bokor [44]. Some serious obstacles remain however.

3.1 Self-calibrating algorithms using the Fourier transform method on each frame

Assuming that the spectral sidelobes of the fringe pattern described by Eq. (1) can be perfectly isolated, then application of the FTM [17] yields

$$b(x, y)\exp\left[i\psi(x, y)+i\delta_n\right], \tag{3}$$

so that simply subtracting the phases from each frame gives the relative phase-shift directly. Unfortunately the quantity in Eq. (3) imay not be possible to estimate in general for a three main reasons:

i.) Discontinuities in the spatial fringe patterns cause spectral sidelobes to expand into the entire frequency domain,

ii.) The finite size of the spectral mask used to isolate sidelobes causes artefacts (leakage) in the spatial domain,

iii.) Non-monotonicity of phase $\psi(x, y)$, for closed fringes introduces sign ambiguities in phase estimate.

But problems i) and ii) have recently been solved with a new and truly isotropic 2-D Hilbert-Fourier demodulation algorithm (for brevity referred to as the vortex transform from hereon) developed by Larkin, Bone and Oldfield [45]. Problem iii) resolves itself when more than 2 frames are processed. In the remainder of this paper the vortex transform is used to determine the phase difference between pairs of fringe patterns and hence to define a general auto-calibrating algorithm. The method presented in the following sections, although firmly based on the work of Lai and Yatagai, and Kreis, introduces several important improvements. The method may be near optimal in terms of using all the available intra-frame data for estimating inter-frame phase differences. Proof of optimality may require statistical methods such as Fisher information [34], and will not be pursued here. The method also works with three or more frames, whereas most other methods require at least five frames.

4. Inter-frame phase difference estimation using the vortex transform

The starting point is Eq. (1), which is applicable to all N frames of the sequence. It is well known that the fringe offset term $a(x,y)$ can seriously limit the applicability of the FTM. The

same is true for the vortex transform, so the problem is sidestepped by first forming the inter-frame intensity difference

$$g_{nm} = f_n - f_m = -g_{mn} = b\{\cos[\psi + \delta_n] - \cos[\psi + \delta_m]\} \qquad (4)$$

The function g_{nm} is a pure AM-FM (amplitude modulation and frequency modulation) function without offset and so can be demodulated by the following process if the fringe pattern is locally simple. Locally simple [46] essentially means that the fringe pattern has a unique orientation $\beta(x,y)$ and spacing $\lambda(x,y)$ at each location (x,y) (except, perhaps, at a finite number of locations with uncertain orientation and/or spacing). Local simplicity is the special property which distinguishes fringe patterns from other, more general, patterns.

The spiral phase Fourier operator $\${\}$ [45] is defined as follows:

$$\${f(x,y)\} = F^{-1}\{\exp[i\phi]F\{f(x,y)\}\}. \qquad (5)$$

Wherein the Fourier and inverse Fourier transforms are defined by

$$\left.\begin{array}{l} F\{f(x,y)\} = F(u,v) = \displaystyle\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\exp[-2\pi i(ux+vy)]\,dxdy \\[4mm] F^{-1}\{F(u,v)\} = f(x,y) = \displaystyle\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} F(u,v)\exp[+2\pi i(ux+vy)]\,dudv \end{array}\right\}. \qquad (6)$$

The spectral polar coordinates are given by

$$\left.\begin{array}{l} u = q\cos\phi \\ v = q\sin\phi \end{array}\right\}. \qquad (7)$$

The Fourier operators $F\{\}$ and $F^{-1}\{\}$ transform a general function from spatial domain (x,y) to spatial frequency domain (u,v). The spiral phase $\exp[i\phi]$ is purely a function of the angular component of the spatial frequency polar coordinates (q,ϕ). Applying the spiral phase operator to the inter-frame difference gives the approximation

$$\${g_{nm}\} \cong i\exp[i\beta]b[\sin(\psi + \delta_n) - \sin(\psi + \delta_m)] \qquad (8)$$

The spiral phase operator is an isotropic quadrature operator in that it converts cosine fringes at any orientation into sine fringes at the same orientation, and can be considered as a 2D Hilbert transform operator. The above approximation is good except in regions where the modulation, orientation or spacing varies too rapidly (over the space of one fringe). Areas where the fringe radius of curvature is smaller than the fringe spacing also give a lower accuracy approximation. The full analysis (using the asymptotic method of stationary phase) appears in a recent paper and thesis by Larkin [47, 48]. However, for most typical fringe patterns the approximation is very good and any errors are isotropically distributed (unlike the highly directional errors in more conventional methods such as the half-plane Hilbert transform [21], or spectral sidelobe masking [49]).

Ideally the explicit dependence of Eq. (8) on the fringe orientation $\beta(x,y)$ should be removed because is superfluous to our calibration needs. To do this the orientation must be estimated A number of ways to do this are given in the appendix of Larkin's asymptotic proof [47] of

Eq. (8). One method relies on the gradient of the fringe pattern. The ratio of the x and y components gives the tangent angle from the local Taylor's series expansion of the phase $\psi(x,y)$ about the point (x_0,y_0):

$$\psi(x,y) = \psi_{00} + 2\pi u_0(x-x_0) + 2\pi v_0(y-y_0) + ... \tag{9}$$

The local phase gradient components are

$$\left.\frac{\partial\psi(x,y)}{\partial x}\right|_{\substack{x=x_0\\y=y_0}} \cong 2\pi u_0 \atop \left.\frac{\partial\psi(x,y)}{\partial y}\right|_{\substack{x=x_0\\y=y_0}} \cong 2\pi v_0 \Bigg\} . \tag{10}$$

If the modulation is slowly varying, then the gradient components are

$$\left.\begin{array}{l}\dfrac{\partial g_{nm}}{\partial x} \cong b\left(-\sin(\psi+\delta_n)+\sin(\psi+\delta_m)\right)\dfrac{\partial\psi}{\partial x}\\[2mm] \dfrac{\partial g_{nm}}{\partial y} \cong b\left(-\sin(\psi+\delta_n)+\sin(\psi+\delta_m)\right)\dfrac{\partial\psi}{\partial y}\end{array}\right\} \tag{11}$$

The ratio of the gradient components gives the tangent

$$\frac{\partial g_{nm}}{\partial y} \bigg/ \frac{\partial g_{nm}}{\partial x} \cong \frac{v_0}{u_0} = \tan\beta_{nm} \tag{12}$$

Again the approximation is valid in regions with smoothly varying parameters as defined for the spiral phase operator. Other methods of orientation estimation may be used. It is even possible to use the spiral phase operator itself for orientation estimation, as Eq. (8) contains the orientation information in its phase. However the overall process becomes more involved, so for clarity a generic orientation estimation is assumed at this point. Clearly, the orientation for any frame f_n or inter-frame difference g_{nm} may be calculated. Alternatively the orientation may be estimated from combinations of the previous methods in a statistically advantageous manner. Assuming just one orientation estimate (however found) with the value β_e, then the estimate typically contains an ambiguity in the sign:

$$\exp[i\beta_e] = \cos\beta_e + i\sin\beta_e = \pm\frac{u_0+iv_0}{\sqrt{u_0^2+v_0^2}} . \tag{13}$$

Different orientation estimates typically have the sign ambiguities in different places. It is worth noting that the overall sign ambiguity arises from the general difficulty in distinguishing a fringe at orientation β from a fringe at $\beta+\pi$. Rewriting the inter-frame differences:

$$g_{nm} = 2b\sin\left[\psi+(\delta_n+\delta_m)/2\right]\sin\left[(\delta_n-\delta_m)/2\right] \tag{14}$$

then the vortex operator V{ } counteracts the orientation phase:

$$\mathrm{V}\{g_{nm}\} = -i\exp[-i\beta_e]\$\{g_{nm}\} \cong 2b\exp[i(\beta - \beta_e)]\cos[\psi + (\delta_n + \delta_m)/2]\sin[(\delta_n - \delta_m)/2]$$

(15)

The complex exponential on the RHS above is essentially ±1, depending on the orientation estimate. It is convenient to call this the polarity factor *h(x,y)* [50]:

$$h(x,y) = \exp\left[i(\beta - \beta_e)\right] = \begin{cases} +\exp[i\varepsilon], & \beta_e = \beta - \varepsilon \\ -\exp[i\varepsilon], & \beta_e = \beta - \varepsilon + \pi \end{cases}$$

(16)

The error in the orientation estimate ε introduces a small imaginary deviation from $h=\pm1$. Now combining the two real components from Eq. (15) and Eq. (14) in the real and imaginary parts of a complex function the *contingent analytic image,* $\tilde{g}_{nm}$ is obtained:

$$\tilde{g}_{nm} = \mathrm{V}\{g_{nm}\} + ig_{nm}$$

(17)

$$\tilde{g}_{nm} = 2b\sin[(\delta_n - \delta_m)/2](h\cos[\psi + (\delta_n + \delta_m)/2] + i\sin[\psi + (\delta_n + \delta_m)/2])$$
$$= 2b\sin[(\delta_n - \delta_m)/2]\exp\{ih\cos[\psi + (\delta_n + \delta_m)/2] + i\pi/2\}.$$

(18)

This definition of *analytic image* is analogous to the one-dimensional definition originated by Gabor [51], but is *contingent* upon the polarity of the orientation function β_e used in the definition. In a sense the analytic image is the demodulated image. This definition is consistent with definitions of the analyticity used by Havilicek [52], and Bülow [53], but not the analyticity of Fiddy [54] and other researchers in phase-retrieval. The phase can be calculated for every inter-frame difference as follows:

$$\alpha_{nm} = Arg\left(\tilde{g}_{nm}\right) \cong \left[\psi + (\delta_n + \delta_m)/2\right]h + \pi/2$$

(19)

The phase calculation is possible as long as *sin[(δn-δm)]≠0*, which only occurs if *(δm-δn) ≠2jπ*, where *j* is an integer. Clearly the failure in such a case is due to the phase shifts being essentially equal and the inter-frame difference being near zero. As expected an inter-frame difference maximum corresponds to fringes in antiphase.

It is useful to consider triplets of frames phases:

$$\alpha_{nm} - \alpha_{mk} = \left[\psi + (\delta_n + \delta_m)/2\right]h - \left[\psi + (\delta_m + \delta_k)/2\right]h = \left[(\delta_n - \delta_k)/2\right]h.$$

(20)

In other words it is possible to find the difference in phase between any frames *n* and *k*, within a global sign ambiguity *h*. Note an implicit assumption is that all phase calculations (including differences and differences of differences) are all calculated modulo 2π.

5. Estimation of phase-shifts using periodic phase statistics over frame areas

The main demodulation task culminates in Eq. (19), while the removal of common phase components occurs in Eq. (20). The phase-shift in Eq. (20) can be evaluated at every pixel of a fringe pattern image. Spatially invariant phase-shifts are assumed in the remainder of this paper, however the possibility of spatially varying phase-shifts can be investigated by taking Eq. (20) and applying the appropriate analysis. Various methods can be used to extract both the sign ambiguity and the phase differences. The initial approach taken in my thesis [48]

utilised the statistics of the phase component alone. So, for example, the absolute values can be calculated and averaged over the full frame area S:

$$\left. \left| \delta_n - \delta_k \right| \right|_{mean} = 2 \frac{\iint_S \left| \alpha_{nm} - \alpha_{mk} \right| dxdy}{\iint_S dxdy} \tag{21}$$

The above equation does not give good results in all situations because phase statistics are intrinsically periodic (modulo 2π). So , for example, if the mean is near $+\pi$, then Eq. (21) introduces severe phase wrapping errors near $-\pi$. Recently there has been renewed interest in the statistics of periodic functions [55]. It transpires that the mean and the variance of the periodic functions are connected. A similar result interlinking the axial and transverse widths of an optical point spread function was derived by Sheppard and Hegedus [56]. An equivalent approach, which is taken here, is to treat all phase calculations as complex valued phasors, as proposed by Strobel [57], in which case the periodic wrapping issue is implicitly avoided. The angular statistics of a distribution localised on the rim of a unit radius disk - like that of Fig. 1 – naturally adhere to periodic restrictions.

Before proceeding further, the global sign ambiguity h_g needs to be fixed for the particular fringe sequence. The relation is:

$$\frac{\alpha_{nm} - \alpha_{mk}}{\left| \alpha_{nm} - \alpha_{mk} \right|} = h(x,y) \frac{\delta_n - \delta_k}{\left| \delta_n - \delta_k \right|} = h(x,y) \operatorname{sgn}\left(\delta_n - \delta_k \right). \tag{22}$$

Arbitrarily taking the α phase difference as positive for the particular values of $\hat{n}$ and $\hat{k}$, then:

$$h_g(x,y) = h_{\hat{n}\hat{k}}(x,y) = \operatorname{sgn}\left(\alpha_{\hat{n}\hat{m}} - \alpha_{\hat{m}\hat{k}} \right), \tag{23}$$

and once defined this particular polarity function estimate can be used in all further computations. The complex (unit modulus) phasor selected for statistical analysis should directly represent the phase-shift of interest, so p_{nk} is defined by

$$p_{nk}(x,y) = \exp\left[2ih\{Arg(\tilde{g}_{nm}) - Arg(\tilde{g}_{mk})\} \right] \cong \exp\left[i(\delta_n - \delta_k) \right]. \tag{24}$$

The right-most expression above is the inter-frame phase difference sought for self-calibration. The middle expression defines a convenient computation procedure for generating the approximation over a full frame. It turns out that the above phase estimate is quite well localised for typical interferograms. The first estimate of phase difference is then calculated from a simple mean (the angle of the centroid of a ring distribution) as follows:

$$\overline{\left(\delta_n - \delta_k \right)}_{mean}^{phase} = Arg\left[\overline{p_{nk}} \right] = Arg\left[\frac{\iint_S p_{nk}(x,y)dxdy}{\iint_S dxdy} \right] = Arg\left[\iint_S p_{nk}(x,y)dxdy \right] \tag{25}$$

A preferable method uses a weighted integral over the frame area. The (real) weighting function $w(x,y)$ used should be a measure of the estimated accuracy of the local phase estimate (this will be related to the AM-FM characteristics later on)

$$\gamma_{nk} = \overline{\left(\delta_n - \delta_k\right)}_{\substack{phase \\ weighted \\ mean}} = Arg\left[\overline{_w p_{nk}}\right] = Arg\left[\frac{\iint_S p_{nk}(x,y)w^2(x,y)dxdy}{\iint_S w^2(x,y)dxdy}\right] \qquad (26)$$

So, for example, a very low (or zero) weight is appropriate where the phase varies too quickly (viz. discontinuities) or the modulation is too low. An initial estimate of the weight function $w(x,y)$, simply based on the contingent analytic image magnitude of Eq. (17), may be used in the first phase estimate.

The next step is to calculate $\delta_n - \delta_k$ for all possible pairs $n \neq k$. The minimum requirement is to calculate the phase shift difference for all sequential pairs, which amounts to N values for N frames. By calculating non-sequential pairs additional statistical robustness may be obtained in the estimates. The total number of two-phase differences combination from N phases being

$$\frac{N!}{2(N-2)!} \qquad (27)$$

In the particularly interesting case of $N=3$, the number of combination is also 3, which equals the number of sequential differences too. Fig. 1 shows all possible inter-frame differences for the case where $N=5$. The length of the connecting line is a direct indication of the magnitude related to the phase difference estimate. In fact the line length is proportional to $sin|(\delta_n-\delta_m)/2|$, a factor which appears in Eq. (18).
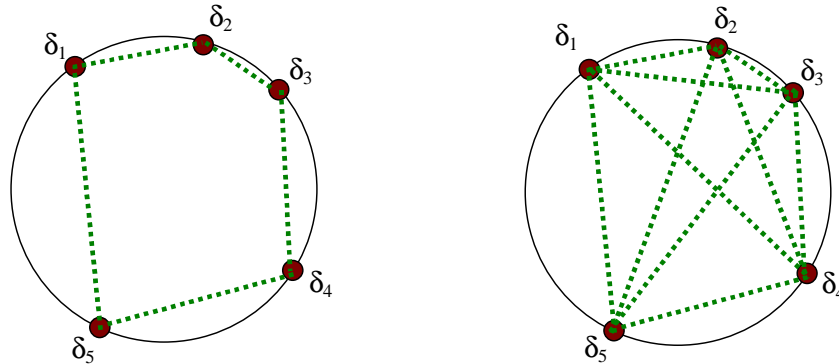

Fig. 1   The left plot shows sequential inter-frame differences (dotted connecting lines) on a unit circle representing phase angles from 0 to 2π. The right plot shows all possible inter-frame differences as dotted lines

As an example consider all ten possible phase difference pairs for N=5 as shown in Fig. 1, where $\gamma_{mn}=\delta_m-\delta_n$:

$$\begin{cases} \gamma_{12}, & \gamma_{23}, & \gamma_{34}, & \gamma_{45}, \\ \gamma_{13}, & \gamma_{24}, & \gamma_{35}, \\ \gamma_{14}, & \gamma_{25}, \\ \gamma_{15}, \end{cases} \qquad (28)$$

In situations where the frames contain noise and other distortions, the best estimates of γ_{mn} may be expected to come from a two-dimensional fit of the all the points (as shown in Fig. 1) to a unit radius circle. The fitting procedure is better tempered than the ellipse fitting

considered in section 2.3 because the unit circle has no free parameters. Details are not pursued here, suffice to say that such methods may be used to beneficial effect when estimating all values of γ_{mn} in this proposed calibration method.

Without loss of generality, one of the phase shifts can be set to zero for convenience, for example $\delta_1 = 0$. This is consistent with the idea that the absolute value of the phase is undefined and only phase differences are really meaningful. The remaining phases can be calculated in sequence

$$\delta_1 = 0, \ \delta_2 = \delta_1 + \gamma_{21}, \ \delta_3 = \delta_2 + \gamma_{32}, \ \delta_4 = \delta_3 + \gamma_{43}, \ \delta_5 = \delta_4 + \gamma_{54} \tag{29}$$

and so on for larger values of N. These values can now be substituted back into the general PSA defined by Morgan [25] and Grievenkamp [26]. and all the fringe parameters (offset, modulation and phase) calculated.

At this stage there may be some residual calibration error due to the original vortex operator failing in certain areas. These errors are rectified by re-evaluating the weight function $w(x,y)$ based on new estimates of $a(x,y)$, $b(x,y)$, $\psi(x,y)$, and their partial derivatives. This allows an estimate of how well the original frames fulfilled the smoothness requirements of the vortex operator [47]. Once the new weight function is evaluated the phase calibration process is repeated (from Eq. (26) onwards) resulting in a better estimate. The accuracy may be characterised by the (periodic function) variance of the phase difference distribution [55]:

$$\sigma_{nk}^2 = \frac{\iint_S |p_{nk} - {}_w\bar{p}_{nk}|^2 w^2(x,y) \, dx \, dy}{\iint_S w^2(x,y) \, dx \, dy}. \tag{30}$$

For the periodic function p_{nk} the variance is simply dependent on the mean (or centroid):

$$\sigma_{nk}^2 = 1 - |{}_w\bar{p}_{nk}|^2. \tag{31}$$

The periodic function (viz. phase) statistics in Eq. (25), (26), (30), and (31) are robust and well-behaved, making them ideal for inclusion in automated algorithms.

6. Algorithm Summary

Fig. 2(a) shows a flowchart of the proposed automatic phase-shift calibration method. Starting from a sequence of frames f_n the inter-frame differences g_{nm} are formed. The next step is to form the contingent analytic image $\tilde{g}_{nm}$ corresponding to g_{nm} as defined earlier. Two full-frame fast Fourier transforms (FFTs) are used for the vortex transform without any windowing or other modifications normally required by Fourier transform methods (although the analysis presented so far is for continuous functions and continuous transforms, the discrete sampled image data used in practice follows the continuous analysis closely because sharp discontinuities in Fourier space are neatly avoided by the use of the vortex transform).

From here the phase difference $\alpha_{nm} - \alpha_{mk}$ (modulo 2π) from two dependent analytic images is extracted. An arbitrary global polarity function $h_g(x,y)$ is then derived from the sign of the phase difference.
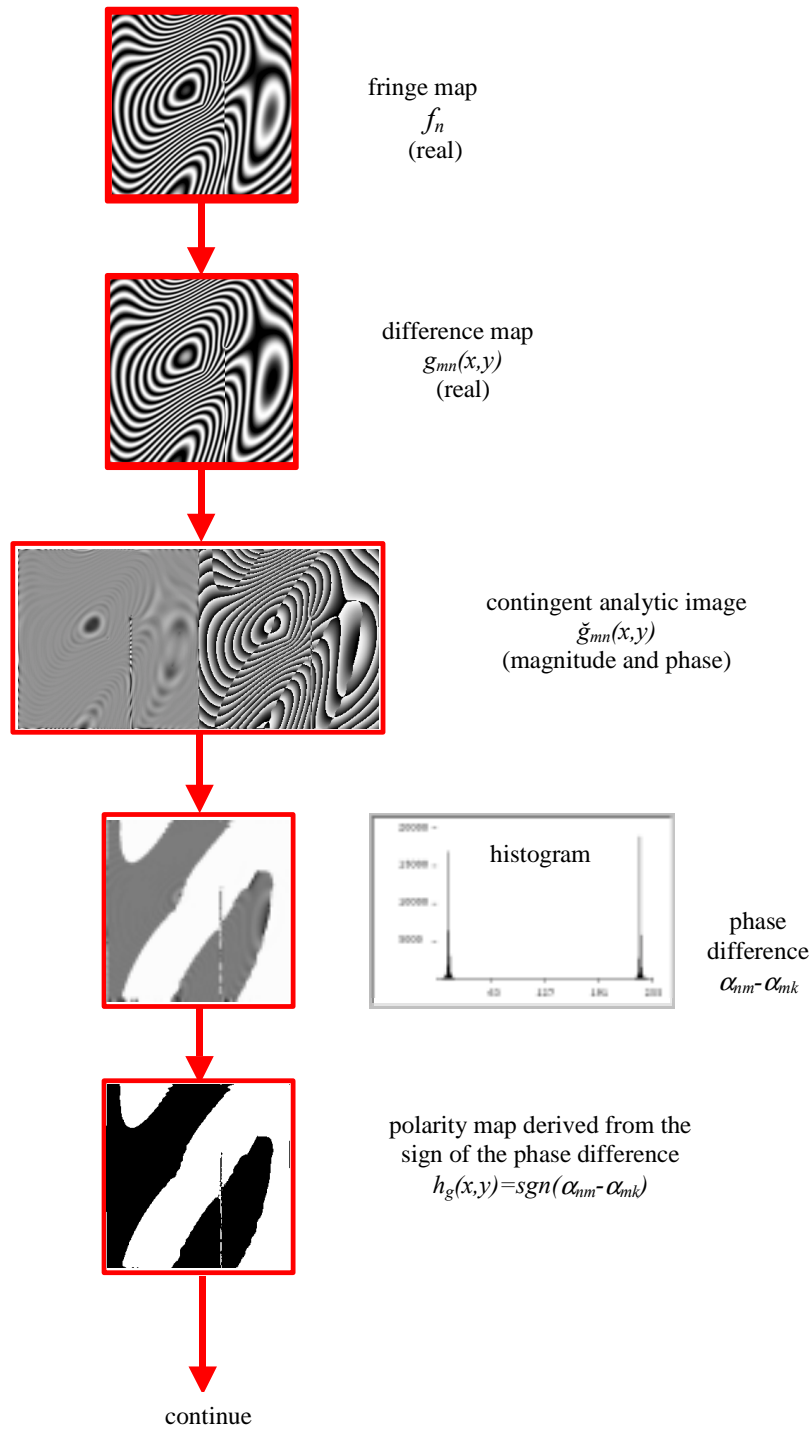
fringe map
f_n
(real)

difference map
$g_{mn}(x,y)$
(real)

contingent analytic image
$\check{g}_{mn}(x,y)$
(magnitude and phase)

histogram

phase
difference
$\alpha_{nm} - \alpha_{mk}$

polarity map derived from the
sign of the phase difference
$h_g(x,y) = sgn(\alpha_{nm} - \alpha_{mk})$

continue

Fig. 2 (a). Flowchart for automatic phase-step calibration method.

polarity map derived from the
sign of the phase difference
$h_g(x,y)=sgn(\alpha_{nm}-\alpha_{mk})$

polarity corrected phase difference
$\delta_n-\delta_k=h_g\{Arg(\breve{g}_{nm})-Arg(\breve{g}_{mk})\}$

then form pure phase function
$p_{nk}=exp[i(\delta_n-\delta_k)]$

$$\delta_1 = 0$$
$$\delta_2 = \delta_1 + \gamma_{21}$$
$$\delta_n = \delta_{n-1} + \gamma_{n(n-1)}$$

estimate phase shifts γ_{nk} utilising
weighted periodic function statistics
(complex phasors)

Using PSA
estimate
modulation b
and phase ψ

$b(x,y)$ $\psi(x,y)$

Using PSA
estimate offset a

$a(x,y)$

No   accurate
result?   Yes

OUTPUT

δ_n, $n=1\rightarrow N$,
$a(x,y)$
$b(x,y)$
$\psi(x,y)$

Fig. 2(b). Continuation of flowchart for automatic calibration method.

In Fig. 2(b) the global polarity function is applied to all pairs of relative phase-shifts in a consistent manner. The pure phase function p_{nk} is used to calculate periodic estimates of the sequential phase shifts. The phase-shift information is then used to generate a generalized PSA for the frame sequence and estimates of the fringe pattern offset $a(x,y)$, modulation $b(x,y)$, and phase $\psi(x,y)$, are generated. At this stage it is possible to characterise the likely errors in the earlier contingent analytic image. The errors are related to the various partial derivatives of the three aforementioned functions, but a simple estimate can be obtained from a direct comparison of the analytic image and the generalised PSA image. Regions with large discrepancies can be give low or zero weight.

Fig. 3 shows a simple example where the above comparison yields a binary weight function, $w(x, y)$ which is 1.0 where the comparison is good (close) and 0.0 where the comparison is bad. This weight function can then be used in a second iteration of the method. The weight function is applied to the estimation of the mean phase-difference where it systematically removes the outliers from the distribution. Finally all the inter-frame phase differences are estimated and a comparison is made of the fringe parameters $a(x, y)$, $b(x, y)$, and $\psi(x, y)$ derived from the vortex transform and the generalised PSA. If the estimates agree within some preset bounds, then the process is terminated and the phase-shift parameters are saved as output data. If the accuracy is insufficient (or the variance in Eq. (30) and (31) is too large) then the process is repeated.
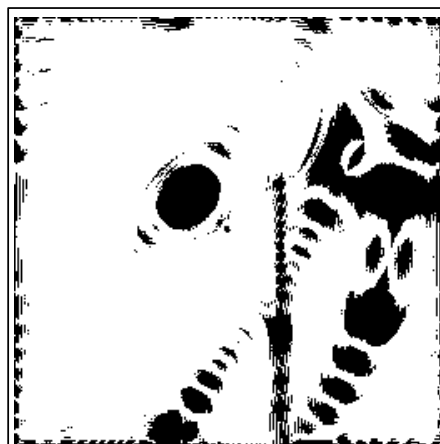

Fig. 3. Simple weight function calculated from the estimated errors in the contingent analytic function (relative to the output from the PSA). Black regions indicate zero weight, white regions have a weight of 1. Note how the weight removes the contribution from regions with large fringe curvature, from regions with discontinuous phase, and from regions with stationary phase.

7. Some Accuracy Estimates

The method presented is not prone to the same errors as the alternative phase-shift calibration methods of section 2. So it is not dependent upon the number of fringes across the frame being large and the fringes linear. It does however assume that the fringes adhere to the requirements of the stationary phase expansion of the spiral phase transform [47]. In that paper the theoretical basis of an error analysis has been outlined so that specific fringe pattern errors can be estimated. Initial experiments with the vortex transform show that it performs

well, even in the presence of substantial additive random noise. Overall systematic errors are expected to be lower than other Fourier methods because leakage and discontinuity artefacts are avoided. Of course the main advantage of the method is that it allows direct estimation of phase shifts for pattern sequences previously considered difficult or impossible. Ideally some standard test patterns should be used to compare methods.

Fig. 4. Fringe pattern used for testing the phase-shift calibration algorithm. Note the closed fringes, the rapid fringe spacing variation and the fringe discontinuity – factors which defeat many other techniques. The sampled intensity varies between 0 and 255 grey levels.

As an initial test the fringe pattern shown in Fig. 2, and again in more detail in Fig. 4, was used. Three fringe images with known phase shifts were generated with 256x256 pixels each of 256 grey-levels. The calibration algorithm was run without a second iteration so the initial accuracy could be tested. The results, shown in table 1, for the error in the final estimated phase are rather good. There is a maximum error of 0.0068 radian, and a standard deviation of 0.0048 which is between 2 and 4 times the underlying error due to intensity quantization [58, 59]. The error in the phase is primarily second harmonic as expected. Fig. 5 shows the actual phase error for the test . The error is less than half that of the min-max method of Chen et al in section 2.5, which quotes a maximum error of 0.015 radian in their best case (no added noise), and requires 15 frames [39]. Direct comparisons with other methods in sections 2.1, 2.2, 2.3, 2.4 and 2.6 are not possible because these method fail for closed fringes or require more than 4 fringe patterns.

Table 1. Performance of phase-shift calibrating algorithm applied to fringe pattern of Fig. 4

| Parameter	Exact value	Valued estimated by algorithm
a	127.5 grey levels	Standard deviation 1.98
b	127.5 grey levels	Standard deviation 0.62, *Quantization 0.29*
Phase error	0.0 by definition	Standard deviation 0.0048 radian Max Min ±0.0068 radian, *Quantization 0.0012 to 0.0027 (uncorrelated, correlated)* [58, 59].
Phase step 1-2	-1.3333	-1.3186 (+0.0147) radian
Phase step 2-3	+2.0404	+2.0174 (-0.0230) radian
Phase step 3-4	-0.7071	-0.6988 (+0.0083) radian
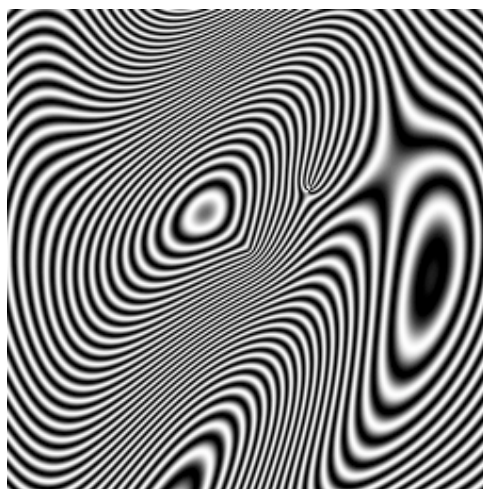
Fig. 5. Phase error showing the classic second harmonic structure, and the disappearance of the vertical half-period discontinuity. Note that the phase error is just ±0.0068 radian, but is shown here normalised.

8. Discussion

A new phase-shift calibration algorithm has been presented and tested on a particularly difficult triplet of fringe patterns containing closed fringes with discontinuities and rapidly varying orientation and spacing. Initial results look very promising. The error levels are already very low and do not rely on large numbers of interferograms for accuracy. The low error levels are believed to be related to the way that the underlying quadrature (vortex) transform utilises the full information from each of the fringe pattern frames. The method has applications in high precision interferometry – where precision mechanical stepping can now be replaced by low precision stepping in conjunction with more sophisticated software. Interferograms with a small number of fringes across the field can now be analysed effectively. The computational burden is not too great, with just two full-frame FFTs required to implement the vortex transform on each frame. Future work will need to address the question of optimality for this self-calibration technique.