5. The Ultimate Phase-Shifting Algorithm?

If one tells the truth, one is sure, sooner or later, to be found out.

Oscar Wilde, *Chameleon*, December 1894.

5.1 A cornucopia of phase-shifting algorithms

We have already met a number of phase-shifting algorithms (PSAs) in chapters 2 and 4. Back in 1980 when the interferometer manufacturer Zygo released its interferogram interpretation handbook[1] a good proportion of the space was given over to the analysis of fringe patterns using ruler and pencil and a chart of fringe patterns from aberration types (see figure 1.1 and 1.2). I can still vividly remember my optics laboratory work (1979) including the photography of Twyman-Green fringe patterns and their subsequent qualitative analysis.[2]

Since that time several hundred papers have been written on the automated digital analysis of fringe patterns, and numerous automated instruments are available commercially. Looking back, the old methods seem rather quaint, although inevitable given the rarity of suitable sensors and digital computers. One of the earliest digital analyses was actually back in 1974, by Bruning et al,[3] but it wasn't until the 1980's that the familiar algorithms[4] had their genesis.

From the beginning automated interferogram analysis split into two broad categories: spatial methods, and temporal (phase-shifting) methods. Initially the

spatial methods were just algorithmic reformulations of the tried and tested pencil and ruler methods. In contrast, the new phase-shifting methods were a major leap in the control and accuracy of interferometric techniques. In retrospect we see that the first PSAs (both self-calibrating) were proposed in the 1960's by Carré, [5] and Rowley and Hamon.[6] One of the enduring problems with temporal PSAs has been the 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 (I shall not attempt a complete listing[5, 7-14]) to compensate for shift errors, originating primarily from the imperfect piezo transducers used to shift the reference mirror. However these algorithms can only compensate deterministic shift errors (such as first, second and third order shift nonlinearities) and non-sinusoidal fringe profiles.[15-17]

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. The most common situation was where just one measurement was possible because of high speed or transient interferometric phenomena. Where "just one measurement" includes measurements which consist of a whole array of simultaneous measurements, namely an image. Initially methods such as fringe maxima (or minima) tracking and fringe skeletonisation closely followed the old manual techniques. Soon it was realised that such methods were far from optimal when the fringe image contained a wide range of intensities. With the advent of CCD image arrays new possibilities appeared because the CCD sensor has a highly linear response to intensity over a huge range (i.e. 6 or more orders of magnitude). In comparison the

old vidicon image sensors were highly nonlinear in response and needed careful calibration to obtain accurate intensity measurements. With highly linear detectors the idea of sinusoid fitting to the interferogram became practical. The advantage with fitting a sinusoid to a fringe profile is that the phase can be estimated at all points – not just the fringe maximum and minimum.[18-21] In fact the idea (using a vidicon) had been proposed in 1972.[22] The fitting of (1-D) phase modulated sinusoids has an elegant Fourier interpretation, first noted by Takeda,[23, 24] then several others.[19, 25, 26] The analysis was formally extended to 2-D by Bone[27] then Roddier.[28] Subsequently the Fourier transform method (or FTM) has been investigated by many researchers, but it has only achieved acceptance in interferometry where significant tilt (or carrier) fringes are tolerable. In systems without dominant carrier fringes, the FTM falls apart because the spectral sidelobes are no longer isolated from each other. Thomas Kreis[29, 30] has investigated the FTM spectral partitioning problem which occurs when the tilt fringes are insufficient or not present at all. The problem can be seen in terms of the overlapping spectral sidelobes of closed curve fringe patterns. Until recently (see chapter 4) the problem has not been solved generally, but rather as a series of ad hoc (and manual) interventions. The manual intervention involves the selection of spectral masks that substantially isolate the spectral sidelobes. When the sidelobes are overlapping rings a sequence of Hilbert half-plane masks have been suggested.[30]

In order to circumvent the problems of the FTM for closed fringes researchers returned to the spatial domain and sinusoid fitting algorithms that parallel the temporal algorithms. These methods now come under the heading of spatial (carrier) phase shifting algorithms.[31] The Zeiss company used such algorithms in their Direct Measuring Interferometer from 1991, but details were not released until 1997.[32] The

method may be viewed as the convolutional approach to the FTM. The problem then becomes one of finding a suitable kernel to demodulate the fringes. Generally the approach has been one dimensional and reliant upon near constant fringe spacing. More recently the use of error-compensating algorithms (originally derived for temporal phase shifting) has widened the range of fringe spacing variation allowed. [33, 34] The method has been extended to 2-D simplistically by allowing evaluation in two perpendicular directions and combing the results.[35]

Thus far spatial and temporal methods have been considered separately. The temporal methods implicitly assume that the phase shifts between frames are known or constrained in some way. So, for example the Carré algorithm assumes that the phase-steps between frames are all equal, but unknown. About ten years ago researchers began considering whether or not the spatial and temporal methods could be combined to relax the restriction upon the phase shift between frames.

5.2 Spatio-temporal algorithms

It transpires, for typical interferograms, that there is significant redundancy in the information contained within a series of phase shifted fringe image frames. Another way of viewing this is as a strong localisation in the spatio-temporal frequency domain;[36] a view considered in later sections.

5.2.1 Fourier method

Lai and Yatagai[37] 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[38, 39] works with 3 or

more frames having arbitrary 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[30] (on page 146) as a method for removing the sign ambiguity in the FTM.

5.2.2 Overdetermined system of linear equations

An alternative approach using the statistical properties of interferogram sequences was proposed by Okada et al[40] 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 n^{th} interferogram frame in a sequence

$$f_n(x,y) = a(x,y) + b(x,y)\cos\left[\psi(x,y) + \delta_n\right]. \tag{5.1}$$

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 (3*M* 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

$$NM \geq 3M + N - 1 \text{ or } N \geq \frac{3M-1}{M-1} \text{ or } M \geq \frac{N-1}{N-3}. \tag{5.2}$$

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[41] proposed a similar method in 1994. Soon after that Han and Kim[42] proposed an iterative least squares fitting technique, followed by Kong and Kim.[43] Perry and McKelvie[44] compared a number of least–squares estimation algorithms, concluding that the initial phase-shift estimates must be within 5º for the final phase to be correct within $\lambda/100$ (3.6º). 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 are 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.[45]

A recent approach to the problem seems to have evolved in complete isolation from the least squares fitting methods above. Rogala and Barrett[46] derive a maximum likelihood estimate through purely statistical methods and show that several well known PSAs are truly optimal.

5.2.3 Lissajou ellipse fitting method

The approach of Farrell and Player[47, 48] is a geometric interpretation of equation (5.1). At any two fixed pixel locations equation (5.1) 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[49] and may be expected to make this approach to phase calibration more attractive.

5.2.4 Interferogram correlation method

Van Brug[50] 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 equation (5.1)) is closely related to the Fourier transform approach implied by Kreis,[30] 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 stationary. 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.

5.2.5 Fringe pattern min-max comparison

Recently Chen et al[51, 52] 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).

5.2.6 Other methods

Recently Schwider et al[53] have proposed detection of the second harmonic phase errors (which are a characteristic of phase step miscalibration[54]). 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. A 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 linear and equally spaced, and therefore not very interesting.

5.3 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[55] 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 phase gradients. Windecker and Tiziani[56] 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? As mentioned in section 5.2.1, 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[30] on page 146), but no subsequent published work seems to have developed the idea.

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

Assuming that the spectral sidelobes of the fringe pattern described by equation (5.1) can be perfectly isolated, then application of the FTM yields

$$b(x,y)\exp\left[i\psi(x,y)+i\delta_n\right], \tag{5.3}$$

so that simply subtracting the phases from each frame gives the relative phase shift directly. Unfortunately the quantity in equation (5.3) is not possible to estimate directly for 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 already been solved with a new and truly 2-D Fourier demodulation algorithm investigated in chapter 4. Problem III resolves itself when more than 2 frames are processed. The method presented in the following sections 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,[46] and will not be pursued here. The method also works with three or more frames, whereas other methods require at least five frames.

5.4 Inter-frame phase difference estimation using the vortex transform

The starting point is equation (5.1), which is applicable to all *N* frames of the sequence. The first step is to form the inter-frame intensity difference

$$g_{nm} = f_n - f_m = -g_{mn} = b\left\{\cos\left[\psi + \delta_n\right] - \cos\left[\psi + \delta_m\right]\right\} \tag{5.4}$$

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[57] 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). In fact local simplicity is the property which distinguishes fringe patterns from other, more general, patterns.

Define a spiral phase Fourier operator V{} (hereinafter referred to as the vortex operator) as follows:

$$\mathrm{V}\left\{f(x,y)\right\} = \mathrm{F}^{-1}\left\{\exp\left[i\phi\right]\mathrm{F}\left\{f(x,y)\right\}\right\}. \tag{5.5}$$

The Fourier and inverse Fourier transforms are defined by

$$\left.\begin{aligned} \mathrm{F}\{f(x,y)\} &= F(u,v) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\exp\left[-2\pi i(ux+vy)\right]dxdy \\ \mathrm{F}^{-1}\{F(u,v)\} &= f(x,y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} F(u,v)\exp\left[+2\pi i(ux+vy)\right]dudv \end{aligned}\right\}. \quad (5.6)$$

The spectral polar coordinates are given by

$$\left.\begin{aligned} u &= q\cos\phi \\ v &= q\sin\phi \end{aligned}\right\} \quad (5.7)$$

The Fourier operators $\mathrm{F}\{\ \}$ and $\mathrm{F}^{-1}\{\ \}$ 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 vortex operator to the inter-frame difference gives the approximation:

$$\mathrm{V}\{g_{mn}\} \cong i\exp[i\beta]b\left[\sin(\psi+\delta_n)-\sin(\psi+\delta_m)\right] \quad (5.8)$$

The vortex operator is an isotropic quadrature operator in that it converts cosine fringes at any orientation into sine fringes at the same orientation. 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 appears in Chapter 4. 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, or spectral sidelobe masking).

The next step involves compensating for the local fringe orientation $\beta(x,y)$. To do this orientation must be estimated. A number of ways to do this are given in the appendix of Chapter 4. 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):

$$\psi(x,y) = \psi_{00} + 2\pi u_0(x-x_0) + 2\pi v_0(y-y_0) + O\left(x^p y^q\right) \tag{5.9}$$

The local phase gradient components are

$$\left.\begin{aligned} \left.\frac{\partial\psi(x,y)}{\partial x}\right|_{\substack{x=x_0\\y=y_0}} &\cong 2\pi u_0 \\ \left.\frac{\partial\psi(x,y)}{\partial y}\right|_{\substack{x=x_0\\y=y_0}} &\cong 2\pi v_0 \end{aligned}\right\}. \tag{5.10}$$

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

$$\left.\begin{aligned} \frac{\partial g_{nm}}{\partial x} &\cong b\left(-\sin(\psi+\delta_n)+\sin(\psi+\delta_m)\right)\frac{\partial\psi}{\partial x} \\ \frac{\partial g_{nm}}{\partial y} &\cong b\left(-\sin(\psi+\delta_n)+\sin(\psi+\delta_m)\right)\frac{\partial\psi}{\partial y} \end{aligned}\right\} \tag{5.11}$$

The ratio of the gradient components gives the tangent

$$\frac{\partial g_{nm}}{\partial y}\Big/\frac{\partial g_{nm}}{\partial x} \cong \frac{v_0}{u_0} = \tan\beta_{nm} \tag{5.12}$$

Again the approximation is valid in regions with smoothly varying parameters as defined for the vortex operator. Other methods of orientation estimation may be used. It is even possible to use the vortex operator itself for orientation estimation, as equation (5.8) contains the orientation information in its phase. However the overall process becomes more involved, so for simplisticity 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. Assume just one orientation estimate (however found) with the value β_e. The estimate typically contains an ambiguity in the sign:

$$\exp[i\beta] = \cos\beta + i\sin\beta = \pm\frac{u_0 + iv_0}{\sqrt{u_0^2 + v_0^2}}. \tag{5.13}$$

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:

$$g_{nm} = 2b\sin\left[\psi + \left(\delta_m + \delta_n\right)/2\right]\sin\left[\left(\delta_m - \delta_n\right)/2\right] \tag{5.14}$$

and

$$\exp\left[-i\beta_e\right]\mathrm{V}\left\{g_{nm}\right\} \cong 2ib\exp\left[i\left(\beta - \beta_e\right)\right]\cos\left[\psi + \left(\delta_m + \delta_n\right)/2\right]\sin\left[\left(\delta_m - \delta_n\right)/2\right] \tag{5.15}$$

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

$$h(x,y) = \exp\left[i(\beta - \beta_e)\right] \tag{5.16}$$

Now combining components in a complex function we obtain what I shall call the *contingent analytic image,* $\tilde{g}_{nm}$:

$$\tilde{g}_{nm} = i\left(g_{nm} - \exp\left[-i\beta_e\right] \mathrm{V}\{g_{nm}\}\right) \tag{5.17}$$

$$\tilde{g}_{nm} \simeq$$

$$2b\sin\left[(\delta_m - \delta_n)/2\right]\left(h\cos\left[\psi + (\delta_m + \delta_n)/2\right] + i\sin\left[\psi + (\delta_m + \delta_n)/2\right]\right) \tag{5.18}$$

This definition of analytic image is contingent upon the polarity of the orientation function β_e used in the definition. This definition is consistent with definitions of the analyticity used by Havilic,[58] and Bülow,[59] but not the analyticity of Fiddy[60] and other researchers in phase-retrieval. Note that the above combination can be achieved by placing the inter-frame difference in the imaginary part of a complex image, and placing the (imaginary) output of the vortex operator in the real part. The phase of resultant can be calculated for every pair of inter-frame differences as follows:

$$\alpha_{nm} = Arg\left(\tilde{g}_{nm}\right) \cong \left[\psi + (\delta_m + \delta_n)/2\right]h \tag{5.19}$$

The phase calculation is possible as long as $\sin\left[\left(\delta_m - \delta_n\right)/2\right] \neq 0$, which only occurs if $\delta_m - \delta_n \neq 2\pi j$ 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 identically zero.

It is useful to consider triplets of frames phases:

$$\alpha_{nm} - \alpha_{mk} = \left[\psi + \left(\delta_m + \delta_n\right)/2\right]h - \left[\psi + \left(\delta_k + \delta_m\right)/2\right]h = \left[\left(\delta_n - \delta_k\right)/2\right]h . \tag{5.20}$$

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. Various methods can be used to extract both the sign ambiguity and the phase differences. In the simplest embodiment the absolute values are calculated and averaged over the full frame area *S*:

$$\left|\delta_n - \delta_k\right|_{mean} = 2\frac{\iint_S \left|\alpha_{nm} - \alpha_{mk}\right| dxdy}{\iint_S dxdy} \tag{5.21}$$

A preferable method uses a weighted integral over the frame area. The weighting *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)

$$\left|\delta_n - \delta_k\right|_{\substack{weighted \\ mean}} = 2\frac{\iint_S \left|\alpha_{nm} - \alpha_{mk}\right| w^2\left(x,y\right) dxdy}{\iint_S w^2\left(x,y\right) dxdy} . \tag{5.22}$$

So, for example, a very low (or zero) weight is appropriate where the phase varies too quickly (discontinuities) or the modulation is too low. The global sign ambiguity *h* is then simply defined from:

$$\frac{\alpha_{nm}-\alpha_{mk}}{\left|\alpha_{nm}-\alpha_{mk}\right|}=h(x,y)\frac{\delta_n-\delta_k}{\left|\delta_n-\delta_k\right|} \tag{5.23}$$

Arbitrarily taking the phase difference as positive for these particular values of n and k, then:

$$h(x,y)=\frac{\alpha_{nm}-\alpha_{mk}}{\left|\alpha_{nm}-\alpha_{mk}\right|} \tag{5.24}$$

and this polarity function can be used in all further computations because a global sign convention has effectively been set for the phase-shifts in the frame sequence.

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

$$\frac{N!}{2(N-2)!} \tag{5.25}$$

In the particularly interesting case where *N*=3, the number of combination is also 3, which equals the number of sequential differences too. Figure 5.1 shows all possible inter-frame differences for the case where *N*=5. The length of the connecting line is a

direct indication of the intensity of the phase difference estimate. In fact the line length is proportional to $\sin\left[\left(\delta_m - \delta_n\right)/2\right]$, a factor which appeared in equation (5.18).

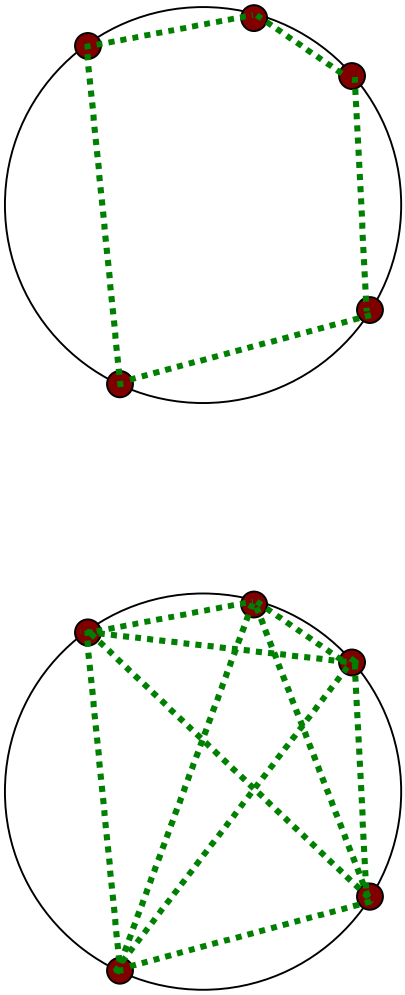

Figure 5.1
Upper plot shows sequential inter-frame differences (dotted connecting lines) on a unit circle representing phase angle. Lower plot shows all possible inter-frame differences as dotted lines

As an example consider the possible phase difference pairs for N=5 as shown in Figure 5.1, where $\gamma_{mn} = \delta_m - \delta_n$.

$$\left.\begin{matrix}\gamma_{12} & \gamma_{23} & \gamma_{34} & \gamma_{45} \\ \gamma_{13} & \gamma_{24} & \gamma_{35} & \\ \gamma_{14} & \gamma_{25} & & \\ \gamma_{15} & & & \end{matrix}\right\} . \tag{5.26}$$

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 5.1) to a unit radius circle. The fitting procedure is better tempered than the ellipse fitting considered in an earlier section because the 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

$$\left.\begin{aligned} \delta_1 &= 0 \\ \delta_2 &= \delta_1 + \gamma_{21} \\ \delta_3 &= \delta_2 + \gamma_{32} \\ \delta_4 &= \delta_3 + \gamma_{43} \\ \delta_5 &= \delta_4 + \gamma_{54} \end{aligned}\right\} \tag{5.27}$$

and so on for larger values of N. These values can now be substituted back into the general PSA defined by Morgan[38] and Grievenkamp.[39] and all the fringe parameters (offset, modulation and phase) calculated.

At this stage there may be some residual calibration error due to our 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. Once the new weight function is evaluated the phase calibration process is repeated. (from equation (5.22) onwards) resulting in a better estimate. The accuracy may be characterised by the variance of the phase difference

$$\sigma^2_{\delta_n-\delta_k} = 4\frac{\iint_S \left(\left|\alpha_{nm}-\alpha_{mk}\right|-\bar{\alpha}_{nk}\right)^2 w^2(x,y)\,dxdy}{\iint_S w^2(x,y)\,dxdy} \tag{5.28}$$

where the mean is defined

$$\bar{\alpha}_{nk} = \frac{\iint_S \left|\alpha_{nm}-\alpha_{mk}\right| w^2(x,y)\,dxdy}{\iint_S w^2(x,y)\,dxdy}. \tag{5.29}$$

Some care needs to taken evaluating these expression as the phase is actually periodic and false discontinuities can disrupt the calculations if $\bar{\alpha}_{nk}$ is near 0 or π. Other methods for estimating these parameters that specifically incorporate the cyclic nature may be used.

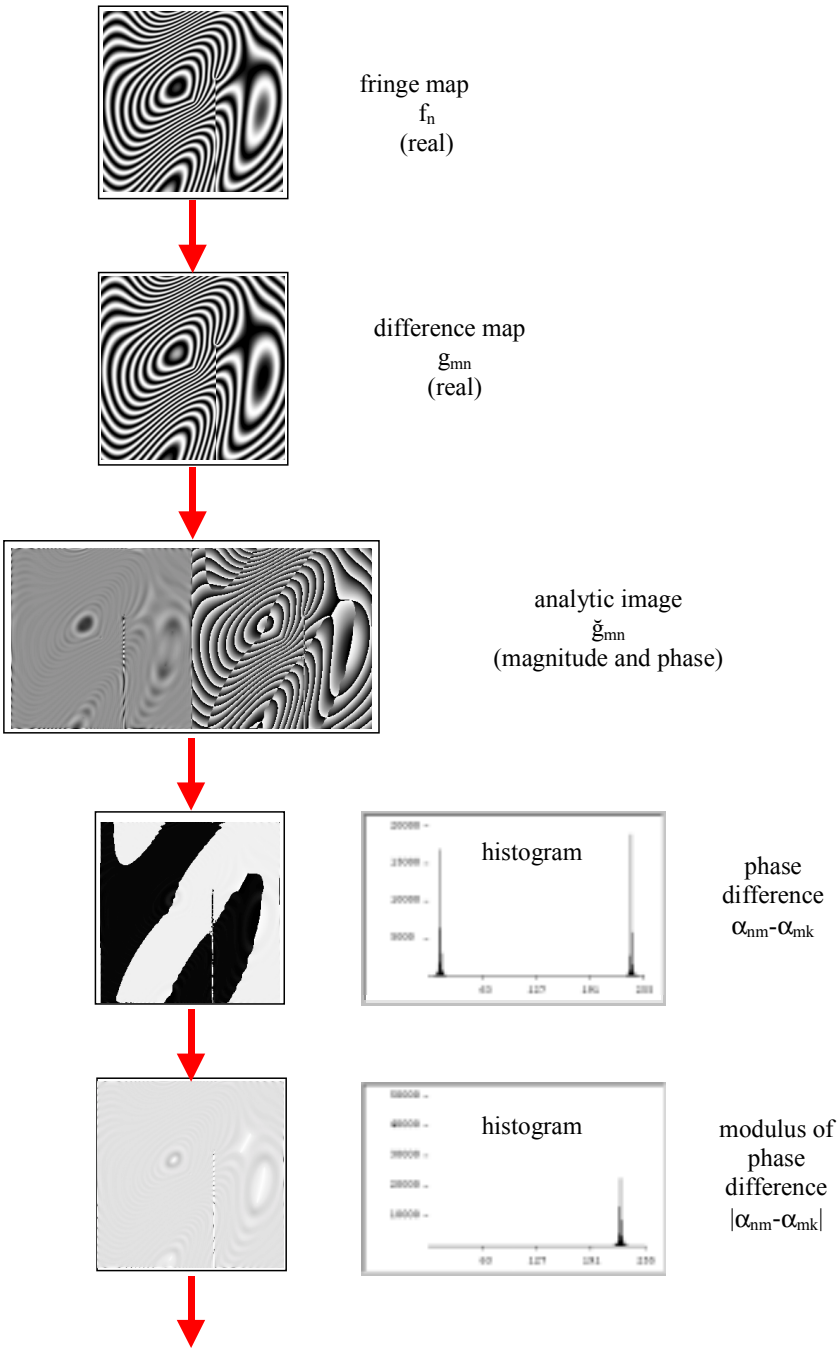

Figure 5.2 (a)
Flowchart for automatic phase-step calibration method.

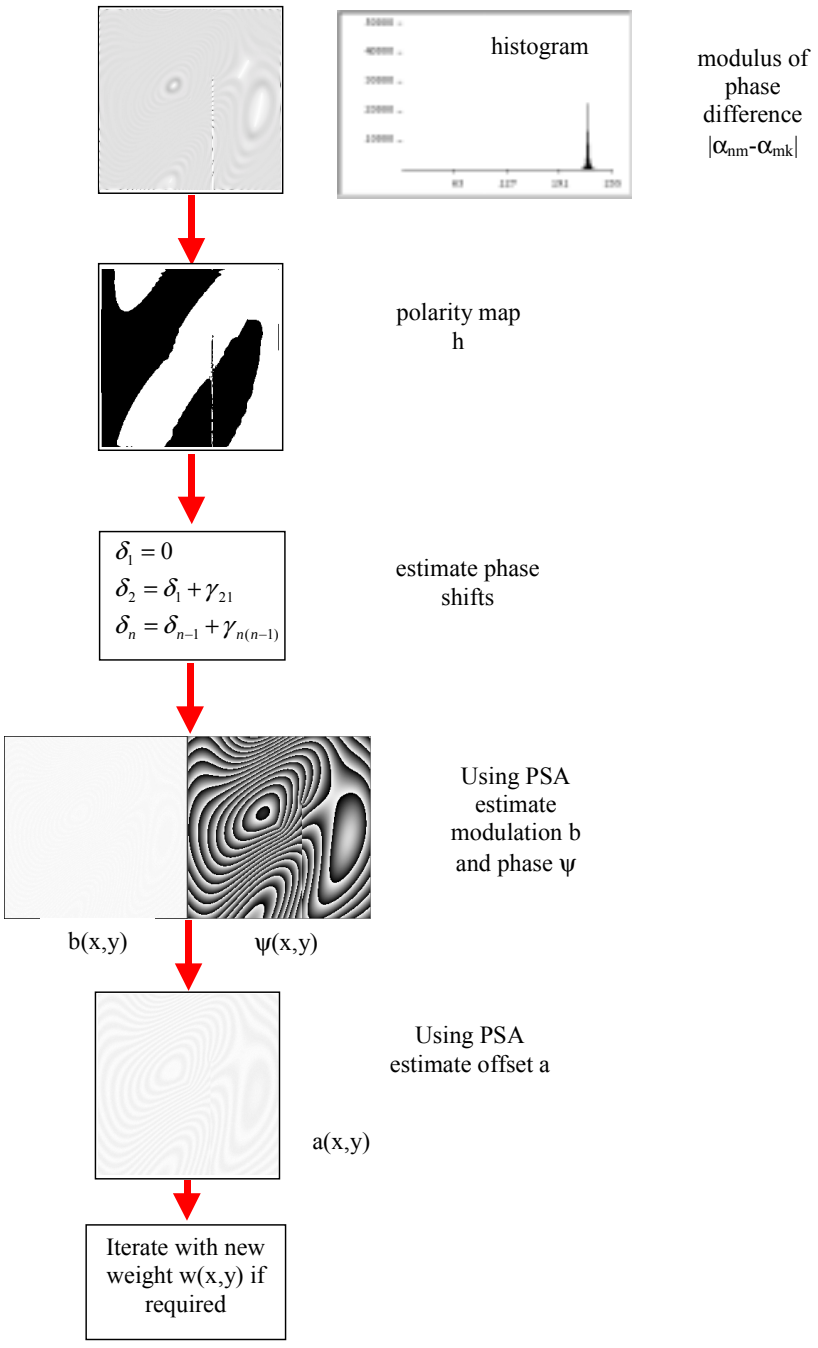

Figure 5.2(b)

Continuation of flowchart for automatic calibration method.

Figure 5.2 shows a flowchart of the 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 FFTs are used for the vortex transform without any windowing or other modifications normally required by Fourier transform methods. From here the phase difference $\alpha_{nm} - \alpha_{mk}$ (modulo 2π) from two dependent analytic images is extracted. Using a particular estimate of the polarity function $h(x, y)$ all pairs of relative phase-shifts are evaluated in a consistent manner. 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 generalized PSA image. Figure 5.3 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. 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 generalized PSA.

Figure 5.3
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.

5.5 Some Accuracy Estimates

The method presented is not prone to the same errors as the alternative phase-shift calibration methods. 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 in Chapter 4. In that chapter the theoretical basis of an error analysis has been laid so that specific fringe pattern errors can be estimated. The response of the method to additive Gaussian noise is not expected to be greatly different to other methods because the quadrature process is expected to generate an independent Gaussian noise pattern. Overall systematic errors are expected to be lower than other methods. Of course the main advantage of the method is that it allows direct estimation of phase shifts for pattern sequences previously considered difficult. Ideally some standard test patterns should be used to compare methods. As an initial test the fringe pattern shown in Figure 5.2,

and again in more detail in figure 5.4, was used Three fringe images with known phase shifts were generated with 256x256 pixels each of 256 greylevels. The calibration algorithm was run without a second iteration so the initial accuracy could be tested. The following results were obtained:

Parameter	Exact value	Value 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)*
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

Table 5.1
Performance of phase-shift calibrating algorithm applied to fringe pattern of Figure 5.4

The results 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. The error is less than half that of the min-max method of Chen et al which quotes a maximum error of 0.015 radian in their best case. The error in the phase is primarily second harmonic as expected. Figure 5.5 shows the actual phase error for the test.

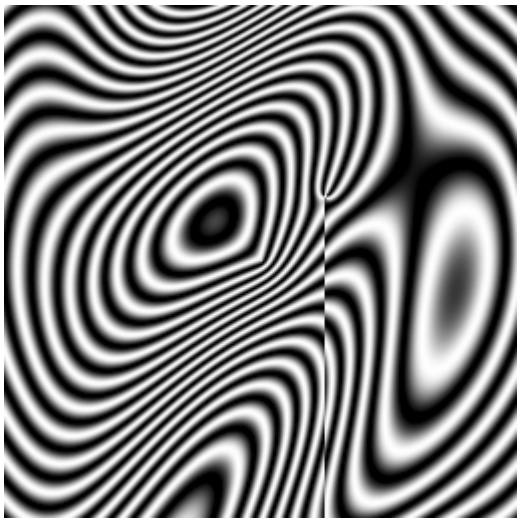

Figure 5.4
Fringe pattern used for testing the phase-shift calibration algorithm. Note the closed fringes, the rapid fringe spacing variation and the fringe discontinuity

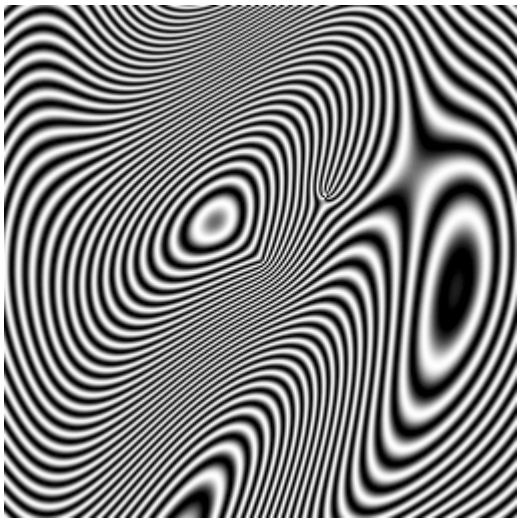

Figure 5.5
Phase error showing the classic second harmonic structure, and the disappearance of the vertical half-period discontinuity. Note that the phase error is small, but is shown here normalised.

5.6 Discussion

A new phase-shift calibration algorithm has been presented and briefly tested on a particularly difficult triplet of fringe patterns. 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 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. The computational burden is not too great, with just two full-frame FFTs required to implement the vortex transform on each frame.

5.7 References and Notes

1 Zygo, "Interferogram Interpretation and Evaluation Handbook", Zygo Corp, (1980).

2 D. Malacara, ed., Optical Shop testing, (New York: John Wiley and Sons, 1978).

3 J. H. Bruning, D. R. Herriot, J. E. Gallagher, D.P.Rosenfeld, et al., "Digital wavefront measuring interferometer for testing optical surfaces and lenses", App. Opt. **13**, 2693-2703, (1974).

4 According to Vladimir Uspensky and Alexie Semenov (in Algorithms: main ideas and applications, Kluwer, Dordercht, 1980) the word "algorithm" is named after al-Khwarizmi (which means from Khorezm) an 8th century scientist. His full name is Muhammad ibn Musa and the Latin transciption of his name (from his treatise on arithmetic) is *alchorismi* or *algorismi*.

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

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

7 Y.-Y. Cheng, and J. C. Wyant, "Phase-shifter calibration in phase-shifting interferometry", App. Opt. **24**, (18), 3049-3052, (1985).

8 P. Hariharan, B. F. Oreb, and T. Eiju, "Digital phase-shifting interferometer: a simple error-compensating phase calculation algorithm", App. Opt. **26**, (13), 2504-2506, (1987).

9 Y. Ishii, and R. Onodera, "Phase-extraction algorithm in laser diode phase-shifting interferometry", Opt. Lett. **20**, (18), 1883-1885, (1995).

10 K. G. Larkin, and B. F. Oreb, "Design and assessment of Symmetrical Phase-Shifting Algorithms", J. Opt. Soc. Am., A **9**, (10), 1740-1748, (1992).

11 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, California, (1992),

12 Y. Surrel, "Phase stepping: a new self-calibrating algorithm.", App. Opt. **32**, (19), 3598-3600, (1993).

13 P. L. Wizinowich, "Phase-shifting interferometry in the presence of vibration: a new algorithm and system", App. Opt. **29**, (22), 3271-3279, (1990).

14 J. Schwider, O. Falkenstorfer, H. Schreiber, A. Zoller, et al., "New compensating four-phase algorithm for phase-shift interferometry", Opt. Eng. **32**, (8), 1883-1885, (1993).

15 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**, (4), 761-768, (1995).

16 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**, (4), 918-930, (1997).

17 K. Hibino, K. G. Larkin, B. F. Oreb, and D. I. Farrant, "Phase-shifting algorithms for nonlinear and spatially nonuniform phase shifts: reply to comment", J. Opt. Soc. Am., A **15**, 1234-1235, (1998).

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

19 W. W. Macy Jr, "Two-dimensional fringe-pattern analysis", App. Opt. **22**, (23), 3898-3901, (1983).

20 K. H. Womack, "Interferometric phase measurement using spatial synchronous detection", Opt. Eng. **23**, (4), 391-395, (1984).

21 P. L. Ransom, and J. V. Kokal, "Interferogram analysis by a modified sinusoid fitting technique", App. Opt. **25**, (22), 4199-4204, (1986).

22 Y. Ichioka, and M. Inuiya, "Direct phase detecting system", App. Opt. **11**, (7), 1507-1514, (1972).

23 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**, (1), 156-160, (1982).

24 M. Takeda, and K. Mutoh, "Fourier transform profilometry for the automatic measurement of 3-D object shapes", App. Opt. **22**, (24), 3977-3982, (1983).

25 K. H. Womack, "Frequency domain desciption of interferogram analysis", Opt. Eng. **23**, (4), 396-400, (1984).

26 K. A. Nugent, "Interferogram analysis using an accurate fully automatic algorithm", App. Opt. **24**, (18), 3101-3105, (1985).

27 D. J. Bone, H.-A. Bachor, and R. J. Sandeman, "Fringe-pattern analysis using a 2-D Fourier transform", App. Opt. **25**, (10), 1653-1660, (1986).

28 C. Roddier, and F. Roddier, "Interferogram analysis using Fourier transform techniques", App. Opt. **26**, (9), 1668-1673, (1987).

29 T. Kreis, "Digital holographic interference-phase measurement using the Fourier transform method", J. Opt. Soc. Am., A **3**, (6), 847-855, (1986).

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

31 M. Kujawinska, "Spatial phase measurement methods", Interferogram analysis: digital fringe pattern measurement techniques, ed. Robinson, D. W., and Reid, G. T. (Bristol: Institute of Physics, 1993).

32 M. F. Kuchel, "Some progress in phase measurement techniques," Fringe97, Bremen, (1997),

33 M. Servin, and F. J. Cuevas, "A novel algorithm for spatial phase-shifting interferometry", J. Mod. Opt. **42**, (9), 1853-1862, (1995).

34 P. H. Chan, P. J. Bryanstoncross, and S. C. Parker, "Fringe-Pattern Analysis Using a Spatial Phase-Stepping Method With Automatic Phase Unwrapping", Measurement Science & Technology **6**, (9), 1250-1259, (1995).

35 M. Pirga, and A. Kujwinska, "Two directional spatial-carrier method for analysis of crossed and closed fringe patterns", Opt. Eng. **34**, (8), 2559-2466, (1995).

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

37 G. Lai, and T. Yatagai, "Generalized phase-shifting interferometry", J. Opt. Soc. Am., A **8**, (5), 822-827, (1991).

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

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

40 K. Okada, A. Sato, and J. Tsujiuchi, "Simultaneous calculation of phase distribution and scanning phase in phase shifting interferometry", Opt. Comm. **84**, (3,4), 118-124, (1991).

41 G. D. Lassahn, J. K. Lassaahn, P. L. Taylor, and V. A. Deason, "Multiphase fringe analysis with unkown phase shifts", Opt. Eng. **33**, (6), 2039-2044, (1994).

42 G.-S. Han, and S.-W. Kim, "Numerical correction of reference phases in phase-shifting interferometry by iterative least-squares fitting", App. Opt. **33**, (31), 7321-7325, (1994).

43 I.-B. Kong, and S.-W. Kim, "General algorithm of phase-shifting interferometry by iterative least-squares fitting", Opt. Eng. **34**, (1), 183-188, (1995).

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

45 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°), even when phase step errors are present..

46 E. W. Rogala, and H. H. Barrett, "Phase-Shifting Interferometry and Maximum-Likelihood Estimation Theory - II - A Generalized Solution", App. Opt. **37**, (31), 7253-7258, (1998).

47 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).

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

49 A. W. Fitzgibbon, M. Pilu, and R. B. Fisher, "Direct least squares fitting of ellipses," International conference on pattern recognition, (1996),

50 H. van Brug, "Phase-step calibration for phase-stepped interferometry", App. Opt. **38**, (16), 3549-3555, (1999).

51 X. Chen, M. Gramaglia, and J. A. Yeazell, "Phase-shifting interferometry with uncalibrated phase shifts", App. Opt. **39**, (4), 585-591, (2000).

52 X. Chen, M. Gramaglia, and J. A. Yeazell, "Phase-shift calibration algorithm for phase-shifting interferometry", J. Opt. Soc. Am., A **17**, (11), 2061-2066, (2000).

53 J. Schwider, T. Dresel, and B. Manzke, "Some considerations of reduction of reference phase error in phase-stepping interferometry", App. Opt. **38**, (4), 655-659, (1999).

54 K. G. Larkin, and B. F. Oreb, "Propagation of errors in different phase-shifting algorithms: a special property of the arctangent function," SPIE Proc. **1755**, San Diego, California, (1992), 219-227.

55 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**, (12), 1439-1444, (1990).

56 R. Windecker, and H. J. Tiziani, "Semispatial, robust, and accurate phase evaluation algorithm", App. Opt. **34**, (31), 7321-7326, (1995).

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

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

59 T. Bülow, and G. Sommer, "A Novel Approach to the 2D Analytic Signal," CAIP'99, Ljubljana, Slovenia, (1999),

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