Maximal fraction of acceptable measurements in phase-shifting speckle interferometry: a theoretical study

Gudmunn Å. Slettemoen*

SINTEF, Norwegian Institute of Technology, N-7034 Trondheim, Norway

James C. Wyant

Optical Sciences Center, University of Arizona, Tucson, Arizona 85721

Received May 20, 1985; accepted August 28, 1985

The interference between a uniform reference wave and a speckle object wave results in variable fringe contrast and background level. Taking these variations into account, we optimize system parameters of phase-shifting speckle interferometry. The results show that the optimal reference intensity should always be equal to one fourth of the detector's saturation level. The optimal reference to the object-intensity ratio shows an increase from one up to, in most practical cases, six as a chosen interference dynamic range increases from its minimum value. The dependence of a maximal fraction of acceptable measurements on the dynamic range is calculated. Numerical examples indicate that we may hope for a readout accuracy in the range of 1/50th to 1/100th of a fringe period and still cover more than half of the image area with acceptable data. These data are taken without spatial averaging and have maximum resolution.

1. INTRODUCTION

Optical phase-shifting techniques are becoming commonly used in interferometric measurements since they provide means of obtaining rapid and accurate phase measurements over the interferogram.[1–3] A major reason for the use of phase-shifting techniques is that measurements can be performed independently of both the variation of fringe contrast and the background level. Since these parameters vary considerably across a speckle pattern, applications of phase-shifting techniques should result in improvements of a variety of speckle interferometric measurements as well.

Holographic interferometry and speckle interferometry usually consist of two separate processes: the recording (primary) process and the reconstruction (secondary) process. In the recording process we register a primary interference pattern ("fringes"). In the reconstruction process we display the secondary fringes, which are a result of combined primary interference fringes. Systems can be classified by the amount of speckle noise that is present. These systems range from classical interferometers that contain no speckle noise to systems that analyze primary speckle interferograms directly. In this paper we shall discuss the latter systems, of which we point out two major examples: those that measure at essentially one point (one speckle) at a time, such as the well-known laser Doppler interferometers,[4] and TV-camera two-dimensional recordings, such as electronic speckle-pattern interferometry (ESPI).[5] ESPI usually displays secondary interferograms, but the primary ones are accessible from the TV camera directly. At the primary stage, the random intensity and phase variations in the speckle pattern have to be taken into account as it seems impossible to average out speckle noise directly, temporally or spatially, at this stage.[6]

In order to apply phase-shifting techniques, we should know how to optimize the setup parameters and how to design calculating/measuring procedures. In this paper we shall quantify these problems by estimating the number of black and saturating spots/speckles in the image. We do this by first calculating the optimum of key parameters of the setups and then finding the fraction of measurements that is considered to be acceptable when a certain minimum measurement accuracy is required. To be more concrete, we may assume that the setup is a specular reference ESPI, although the analysis applies to other specular reference wave setups as well. We denote a wave with uniform intensity as a specular reference wave, as opposed to a speckle reference wave. Previous papers on ESPI and ESPI optimization[6] are all based on the statistical first-order moments of the brightness in the reconstructed image, without considering the complete distribution function. Phase-shifting methods are part of the special case in which self-interference terms are removed from the signal and in which the limited dynamic range alone limits fringe contrast. Having access to the primary interferograms makes it possible to use level discrimination to remove most of the worst data points.

2. PARAMETERS DEFINING ACCEPTANCE LEVELS

We assume that the interference pattern, which is due to interference between a specular (uniform) reference field and a diffusely scattered (speckled) object field, is fully resolved by the detector elements. The intensity pattern is given by the equation

$$I(x, y, t) = |A_r|^2 + |A_0(x, y)|^2 + 2|A_r|\,|A_0(x, y)|\cos[\Phi(x, y) + \Delta\Phi_0(x, y, t) + \Delta\Phi(t)], \quad (1)$$

where

0740-3232/86/020210-05$02.00

A_r is the complex reference field,

$A_0(x, y, t)$ is the complex object field,

$\Phi(x, y)$ is the initial phase difference between the object and the reference field,

$\Delta\Phi_0(x, y, t)$ is the induced phase change in the object field that we want to measure,

$\Delta\Phi(t)$ is the controlled phase shift given by the actual phase-shifting technique being used.

By using the controlled phase shift $\Delta\Phi(t)$ and the intensity measurements $I(x, y, t)$ we want to find the induced phase change $\Delta\Phi_0(x, y, t)$. $\Delta\Phi_0\ (x, y, t)$ can be caused by object movement (deformation/velocity analysis), wavelength change, or directional change of the object-illumination wave (contouring), etc.

To illustrate, let us assume that we want to find $\Delta\Phi_0(x, y, t)$ with intermediate object deformation, as in a holographic double-exposure recording. We then have to find the total relative phase between the reference and the object wave before and after deformation. The difference between these two phase values gives us the desired phase. If a discrete four-step phase-shifting technique is applied directly, we find from Eq. (1) that the phase difference is given by the equation[7]

$$\begin{aligned}\Delta\Phi_0(x, y, t) = {} & \arctan[(I_{3\pi/2}' - I_{\pi/2}')/(I_0' - I_\pi')] \\ & - \arctan[(I_{3\pi/2} - I_{\pi/2})/(I_0 - I_\pi)],\end{aligned} \qquad (2)$$

where I_0, $I_{\pi/2}$, I_π and $I_{3\pi/2}$ are the four intensity measurements corresponding to the four phase settings 0°, 90°, 180°, and 270°, respectively, of $\Delta\Phi(t)$. The unprimed values refer to the intensity levels before deformation, and the primed values refer to the levels after deformation.

In comparison with phase measurements in the secondary interferograms of conventional holography,[8] we have to determine one more parameter, the initial phase difference $\Phi(x, y)$ between the object and the reference wave. However, the direct access to the primary interferograms gives us the freedom of selecting any of the possible combinations between them. But most of all, we get rid of the slow intermediate processes such as the photographic and thermoplastic ones.

In response to the exposure pattern of Eq. (1) we assume that the detector (the TV camera in ESPI) responds linearly within its dynamic range. Since electronic noise will be added to the signal, to apply equations like Eq. (2) we require that

(1) The total intensity always be less than the detector saturation level I_c,

(2) The modulation depth that is due to phase shifting be larger than a certain level $I_{\min}$.

This lower level $I_{\min}$ is given by the phase accuracy that we require from our measurements. Reference 7 discusses how the phase accuracy of classical interferometers depends on the ratio between signal and additive noise and the limited quantization levels in an analog-to-digital conversion. These results are valid for each position x, y in a speckle interferogram as well. To give an example based on four discrete phase steps, we assume that additive electronic noise is limiting our measurement accuracy. Reference 2 shows that if each intensity measurement is made with a signal I (peak modulation depth) to rms noise ratio (I/σ_n), the rms phase error is equal to

$$\Delta\Phi = 1/\sqrt{2}(I/\sigma_n). \qquad (3)$$

In a speckle interferogram this signal-to-noise ratio varies from point to point, and we see from Eq. (1) how the signal varies with the amplitude $|A_0(x, y)|$. If we tolerate a maximum phase error of $\Delta\Phi_{\max}$, we can find from Eq. (3) the minimum value I that is expected to be useful. This value we have already denoted by $I_{\min}$. That is,

$$I_{\min} = \frac{\sigma_n}{\sqrt{2}\Delta\Phi_{\max}}. \qquad (4)$$

We now return to the two general conditions, (1) and (2), which have to be fulfilled before we get acceptable phase-shifted measurements. We have mathematically

$$(1){:}\quad |A_r|^2 + |A_0(x, y)|^2 + 2|A_r|\,|A_0(x, y)| < I_c, \qquad (5)$$

$$(2){:}\quad 2|A_r|\,|A_0(x, y)| > I_{\min} = \frac{I_c}{2D}, \qquad (6)$$

where the useful dynamic range D is defined as the ratio between saturation level I_c and minimum peak-to-peak excursion of the cross-interference term.

By rearranging Eq. (5), we get

$$(1){:}\quad |A_0(x, y)| < (\sqrt{I_c} - |A_r|). \qquad (7)$$

From conditions (6) and (7) we therefore expect the phase-shifting methods to give acceptable data if the amplitude of the diffuse object field fulfills the combined condition

$$\frac{I_c}{4D|A_r|} < |A_0(x, y)| < (\sqrt{I_c} - |A_r|). \qquad (8)$$

In this expression the acceptance levels are defined by the three parameters $|A_r|$, the field amplitude of the uniform reference field; I_c, the saturation level of the detector; and D, which is the dynamic range of the cross-interference signal variations. This range, D, can be related to the phase accuracy as shown by Eq. (4) and expression (6).

3. FRACTION OF ACCEPTABLE MEASUREMENTS

In a specular reference wave setup, only $|A_0|$ is a random function denoting the field amplitude of the speckle pattern. To find the fraction of measurements that fall within the limits of expression (8) we first have to find the probability-distribution function of $|A_0|$. From textbooks on speckle statistics (see, e.g., Ref. 9) we know that in most practical cases the distribution function of the object intensity $|A_0|^2$ is given by the well-known negative exponential function, that is,

$$P_I(|A_0|^2) = \frac{1}{\langle I_0\rangle}\exp\left[-\frac{|A_0|^2}{\langle I_0\rangle}\right], \qquad (9)$$

where $\langle I_0\rangle$ is the statistical average of the intensity $|A_0|^2$.

The distribution of object amplitudes $|A_0|$ can be derived from Eq. (9) by a direct-probability transformation, as given by, e.g., Ref. 10, to give

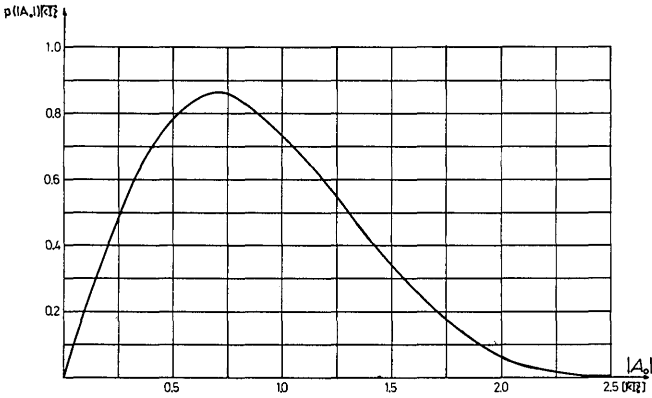

Fig. 1. This graph shows the probability of the speckle field amplitude. The unit of the amplitude scale is equal to the square root of the expectation value of the speckle intensity.

$$p(|A_0|) = \frac{2|A_0|}{\langle I_0 \rangle} \exp\left[-\frac{|A_0|^2}{\langle I_0 \rangle}\right]. \tag{10}$$

By direct inspection we see that this function is, apart from a trivial constant, equal to the derivative of the positive part of the Gaussian distribution function, also called the normal function. It is worthwhile to point out that this function shows that field amplitudes close to zero are relatively rare. A graph of the function is shown in Fig. 1. By integrating between the limits given by expression (8) we get the fraction F of acceptable measurements expressed by the integral

$$F(A_r, I_c, D) = \int_{\frac{I_c}{4D|A_r|}}^{(\sqrt{I_c}-|A_r|)} \frac{2x}{\langle I_0 \rangle} \exp\left[-\frac{x^2}{\langle I_0 \rangle}\right] \mathrm{d}x. \tag{11}$$

By integrating we get the result

$$F(A_r, I_c, D) = \exp\left[-\frac{I_c^2}{16D^2|A_r|^2\langle I_0 \rangle}\right] - \exp\left[-\frac{(\sqrt{I_c}-|A_r|)^2}{\langle I_0 \rangle}\right]. \tag{12}$$

To recast this result into a more manageable form we introduce the normalized parameters

$$r = \text{reference to object-intensity ratio} = \frac{|A_r|^2}{\langle I_0 \rangle},$$

$$R = \text{normalized reference to saturation-intensity ratio} = 4\frac{|A_r|^2}{I_c}. \tag{13}$$

In classical interferometry, in which both waves are uniform, they should have equal intensity. At the point of constructive interference, the total intensity should be equal to the saturation level. When R in Eqs. (13) equals 1, it represents the optimum reference intensity value that we would expect from a classical interferometer. By inserting Eqs. (13) into Eq. (12) we finally find the fraction of acceptable measurements expressed by the normalized parameters D, r, and R, as follows:

$$F(r, R, D) = \exp\left[-\frac{r}{D^2R^2}\right] - \exp\left[-r\left(\frac{2}{\sqrt{R}} - 1\right)^2\right]. \tag{14}$$

4. MAXIMAL FRACTION OF ACCEPTABLE MEASUREMENTS

The fraction F in Eq. (14) has a maximum for a certain optimum reference-to-object intensity ratio r, which can be found by differentiation of F, giving

$$r_{\text{opt}} = \frac{1}{(a-b)} \ln\left(\frac{a}{b}\right), \tag{15}$$

where

$$a = \frac{1}{D^2R^2}, \qquad b = \left(\frac{2}{\sqrt{R}} - 1\right)^2.$$

By inserting r_{opt} into Eq. (14) we see that the optimized fraction will be equal to

$$F(R, D) = \exp[-r_{\text{opt}}a] - \exp[-r_{\text{opt}}b], \tag{16}$$

where r_{opt} is given by Eq. (15).

To find the overall maximum of F with respect to both r_{opt} and the normalized reference to saturation-intensity ratio R, having only the dynamic range D as a parameter, we have to take the derivative of $F(R, D)$ with respect to R in Eq. (16). We first substitute Eq. (15) for r_{opt} and then rearrange Eq. (16) to give

$$F(R, D) = \exp\left[\frac{\frac{a}{b}\ln\left(\frac{a}{b}\right)}{\left(\frac{a}{b}-1\right)}\right] - \exp\left[-\frac{\ln\left(\frac{a}{b}\right)}{\left(\frac{a}{b}-1\right)}\right]. \tag{17}$$

By denoting the ratio a/b as c we can take the following derivative to maximize F:

$$\frac{\delta}{\delta R}(F) = \frac{\delta c}{\delta R}\frac{\delta F}{\delta c}. \tag{18}$$

In this equation either $\delta c/\delta R$ or $\delta F/\delta c$ or both have to be zero to maximize F. By insertion it turns out that the condition $\delta c/\delta R$ equals zero is sufficient. Written out, this becomes

$$\frac{\delta c}{\delta R} = \frac{-2D^2R\left(\frac{2}{\sqrt{R}}-1\right)^2 + 2D^2\sqrt{R}\left(\frac{2}{\sqrt{R}}-1\right)}{D^2R^2\left(\frac{2}{\sqrt{R}}-1\right)^2}. \tag{19}$$

With the numerator in this equation equal to zero we find the two roots:

$$R = 1, \qquad R = 4. \tag{20}$$

Here $R = 1$ represents a maximum of F. This is certainly a simple result:

(1) We get an optimum value of R that is equal to 1. This means that the reference intensity in a specular reference-wave setup, independently of the dynamic range D, should always be equal to the optimum value of a classical interferometer in which the object speckle field is replaced by a uniform one, i.e., a reference intensity equal to the saturation level divided by four.

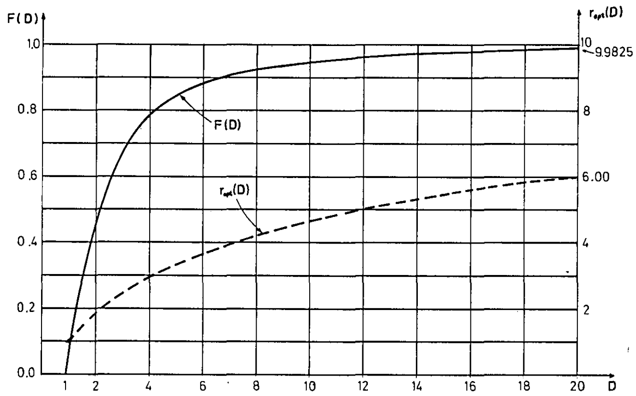

Fig. 2. The optimized fraction $F(D)$ of acceptable data and the optimal reference to object intensity ratio r_{opt} (D) are shown as a function of the dynamic range D. D is an interference dynamic range defined as the ratio between the detector's saturation level and the minimum of the acceptable peak-to-peak excursion of the cross-interference term.

(2) At the same time the optimum reference-to-object intensity ratio from Eq. (15), with $R = 1$ inserted, is equal to

$$r_{opt}(D) = \frac{D^2}{(D^2 - 1)} \ln (D^2). \tag{21}$$

(3) Inserted into Eq. (16), these values give an overall maximum fraction of acceptable measurements that depends on the detectors' dynamic range D only. This maximum is equal to

$$F(D) = \left(1 - \frac{1}{D^2}\right)\exp\left[\frac{\ln(D^2)}{(1 - D^2)}\right]. \tag{22}$$

In the limit $D \Longrightarrow 1$, Eq. (22) shows that the $F(D) \Longrightarrow 0$. This trivial limit shows that if only intensity variations in which the acceptable values have a peak-to-peak value larger than the saturation level, then no measurements are acceptable. As the dynamic range D increases from 1, more and more measurements become acceptable. This is graphically shown in Fig. 2, where the maximal fraction F of Eq. (22) is shown as a function of D. In Fig. 2 we also see how the optimal reference-to-object intensity ratio $r_{opt}(D)$ increases from a minimum of 1 to 6 as D increases from 1 to 20.

5. NUMERICAL EXAMPLES

With classical interferometers we expect to get reliable data from all pixel elements. In a speckle interferometer we deal with a random distribution of intensity causing some of the pixel elements to saturate and others to fail to give fringe modulation above noise level.

Therefore with classical interferometers we may rely on data from neighbor pixel elements also to build up a continuous phase map automatically without 2π ambiguities. On the other hand, with speckle interferometers this is not so straightforward. From what we have calculated, the maximum number of acceptable measurements depends directly on the dynamic range D of the cross-interference signal. As shown by Eq. (4) and expression (6), this range can be related to the required phase accuracy and the detector's dynamic range, which is the ratio between saturation level and rms noise. Explicitly, we have the equation [using Eq. (4) and expression (6)]

$$D = \frac{1}{\sqrt{2}}\left(\frac{I_c}{\sigma_n}\right)\Delta\Phi_{\max}, \tag{23}$$

where the number of discrete phase steps is assumed to be 4.

Let us consider the three examples: a vidicon TV camera with $I_c/\sigma_n = 50$, a linear-detector array camera with $I_c/\sigma_n = 150$, and a photomultiplier tube with $I_c/\sigma_n = 1000$. By use of Eq. (23), the maximum fraction $F(D)$ in Eq. (22) can be written as a function of $\Delta\Phi_{\max}$. Correspondingly, with $I_c/\sigma_n = 50, 150$, and 1000 inserted, the scale of the abscissas in Fig. 2 will be equal to 1.62°, 0.54°, and 0.08° phase accuracy, respectively.

In deformation analysis we find the difference between two phase values, and the resulting phase accuracy will be a factor of $1/\sqrt{2}$ less than the phase accuracy of one phase measurement alone. Therefore these examples show that with the vidicon TV camera we expect no measurements to have an accuracy better than 2.29° (i.e., 1/160th of a fringe). With the linear-detector array and the photomultiplier tube the corresponding numbers would be 0.76° and 0.11°, respectively. From Fig. 2 we see that half of the measurements have an accuracy better than $D = 2.1$, that is, half of the TV camera area will contribute with an accuracy better than $2.1 \times 2.29° = 4.8°$ (i.e., 1/80th of a fringe). With the same TV camera, 90% of the area will be measured with an accuracy of 16° or better.

In order to produce automatically a continuous phase map from a speckle interferogram and still resolve 2π ambiguities, we have to compare phase measurements from neighbor pixel elements. As in classical interferometry,[7] we may require that the phase difference between next-neighbor measurements be less than π. Assume that we want to resolve 2π ambiguities along a line across a TV image or along a linear-detector array. With $\Delta\Phi_{\max} = \pi$ and $I_c/\sigma_n = 150$ inserted into Eqs. (22) and (23), we find that the 2π ambiguities fail to be resolved in only 1 of 7000 trials.

These examples, therefore, indicate that the 2π ambiguity, the fringe-order count, can generally be solved by accepting all data within the maximum dynamic range D, which corresponds to the maximum allowed phase inaccuracy of π (i.e., half a fringe). With a large majority of 2π ambiguities resolved, we can then repeat the procedure with a smaller dynamic range D by selecting other acceptance levels (see Section 2). Thereby the most inaccurate measurements are rejected and better phase accuracy is obtained for the remaining ones. Consequently, by using the results of the initial less accurate measurements, the 2π ambiguities are resolved for the large majority of the most accurate ones.

6. DISCUSSION AND CONCLUSION

In order to make fast and efficient instrumentation of holographic/interferometric setups, we should consider analyzing the interference pattern directly without an intermediate reconstruction step. In ESPI these primary interferograms are coming directly from the TV camera, and by means of a digital frame grabber they are also present in electronic and digital form. However, to analyze a two-dimensional speckle interferogram, we require that phase values across the image be kept track of, a problem that, according to our knowledge, has not been experimentally solved. In order to get rid of 2π ambiguities of induced phase values, the randomness of the phase in the speckle field has to be dealt with. The numerical examples in Section 5 indicate that we may both hope for an accuracy in the range of 1/50th to 1/100th of a fringe and still cover more than half of the image area with acceptable data. Although we may always count fringe orders by direct observation, the the numerical examples also indicate that fringe-order count, the 2π ambiguity, can automatically be taken care of.

In the research reported in this paper we have, by first optimizing the setup parameters, calculated the maximal fraction of measurements that we regard as being acceptable. These results show that the dynamic range, which may also include a limited number of quantization levels,[7] is the key parameter for obtaining a large fraction of measurements in a speckle interferometer. Point detectors, such as P–I–N diodes and photomultiplier tubes, usually have a large dynamic range. In systems depending on point detectors, we consequently seldom measure at positions where data have to be rejected. For example, with a dynamic range (saturation level divided by rms noise) of 1000 and worst accuracy corresponding to 1/100th of a fringe, only 0.5% of randomly positioned detectors will fail. Consequently, with a backup of only one parallel detector, the possibility that both detectors will fail at the same time is only 2.5×10^{-3}%. (The probability that n parallel detectors fail at the same time goes as the nth power of the probability that one of them will fail.)

Finally, we point out that we have assumed the speckle pattern to be fully resolved. This is not the case in practice, and we expect the real fraction of acceptable measurements to be slightly less than the maximum that we have calculated, because the effective dynamic range then decreases. However, it is also worthwhile to point out the generality of our results: independent of the required phase accuracy, and as in classical interferometry, the maximum fraction of acceptable measurements is obtained with a reference intensity that is equal to one fourth of the saturation level. The optimal reference-to-object intensity ratio r_{opt} increases with the dynamic range and is always larger than one.

* Present address, Conoptica a/s, Tonstadgrenda 252, N-7075 Tiller, Norway.

REFERENCES

1. P. Carre, "Installation et utilisation du comparateur photoelectrique et interferentiel du Bureau International des Poids et Mesures," Metrologia **2**, 13 (1966).
2. J. C. Wyant, "Use of an ac heterodyne lateral shear interferometer with real time wavefront correction systems," Appl. Opt. **14**, 2622 (1975).
3. J. H. Bruning, in *Optical Shop Testing*, D. Malacara, ed. (Wiley, New York, 1978).
4. J. B. Abbiss, T. W. Chubb, and E. R. Pike, Opt. Laser Technol. **6**, 249 (1974).
5. R. Jones and C. Wykes, *Holographic and Speckle Interferometry* (Cambridge U. Press, Cambridge, 1983).
6. G. Å. Slettemoen, "First-order statistics of displayed speckle patterns in electronic speckle pattern interferometry," J. Opt. Soc. Am. **71**, 474 (1981).
7. C. L. Koliopoulos, "Interferometric optical phase measurement techniques," Ph.D. dissertation (University of Arizona, Tucson, Ariz., 1981).
8. J. C. Wyant, B. F. Oreb, and P. Hariharan, "Testing aspherics using two-wavelength holography: application of digital electronic techniques," Appl. Opt. **23**, 4020 (1984).
9. J. W. Goodman, *Laser Speckle and Related Phenomena*, C. Dainty, ed. (Springer-Verlag, Berlin, 1975).
10. W. B. Davenport and W. L. Root, *An Introduction to Theory of Random Signals and Noise* (McGraw-Hill, New York, 1958).)