Chapter 8 Three-wavelength single-shot interferometry

8.1 Background and objectives of the study

In the previous chapter, a two-wavelength single-shot method was realized by using a two-wavelength simultaneous imaging system, and a 350 nm step sample was successfully measured. In this method, only the B and R signals of the color camera were used to avoid the crosstalk among the three-wavelength signals. The equivalent wavelength of the two wavelengths used is 1877 nm, and the maximum step that can be measured is about 470 nm.

We studied to extended this method to achieve a three-wavelength single-shot method with the addition of green LED illumination (central wavelength of 530 nm) and the use of camera G-signal for further range expansion [Kitagawa 2008a, Kitagawa 2008b, Kitagawa 2010]. In order to realize this method, new issues such as crosstalk compensation between RGB signals, high-precision frequency estimation, and three-wavelength unwrapping have to be solved. The final flowchart obtained is shown in Fig. 8-1. In this chapter, we describe the three-wavelength imaging system and then describe the details of each algorithm following the flowchart.

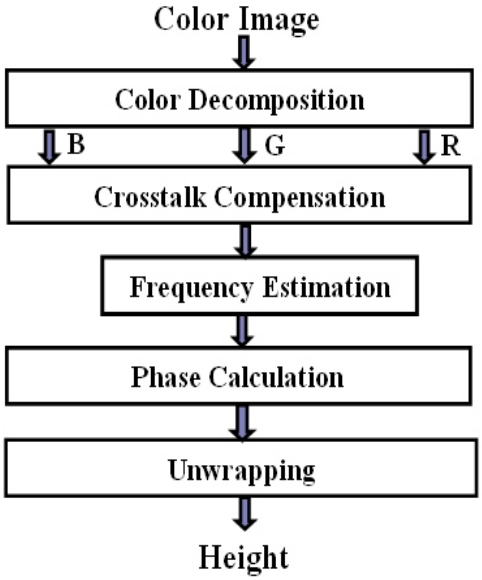


Fig. 8-1 Flowchart of three-wavelength single-shot interferometry [1)]

8.2 Three-wavelength imaging system

The imaging system consists of a commercial three-color LED lighting system and a color camera, as well as the two-wavelength imaging system described in Section 7.3. The center wavelengths (catalog values) of LEDs are 470nm, 530nm, and 627nm.

---

1) In this flowchart, the small block of frequency estimation means that this process is optional in many cases. In other words, the interference fringe frequency estimated here is the carrier fringe frequency determined by the wavelength and reference surface inclination, and does not need to be estimated at each measurement unless the optical system is changed. Even when we perform this estimation, we only need to estimate the entire image once, instead of estimating at each pixel.

Figure 8-2 shows the spectral response characteristics of the color camera used and the central wavelength of the three LEDs. When these three LEDs are simultaneously turned on, a three-wavelength interference image can be obtained by a color camera.

An interference image of a 1 μm standard step sample taken by this system are shown in Figure 8-3. There is a concave step at the bottom center. Decomposition of this image into each color component gave us Fig. 8-4. It is observed that the fringe frequency varies depending on the wavelength. These images are used in the experiments below.

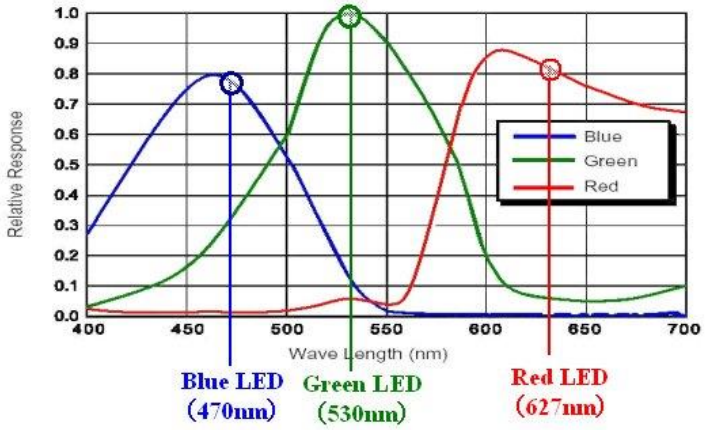


Fig. 8-2 Spectral curves of color camera, with the peak wavelengths of three LEDs

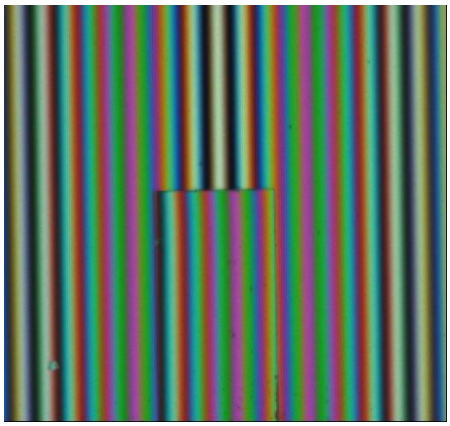
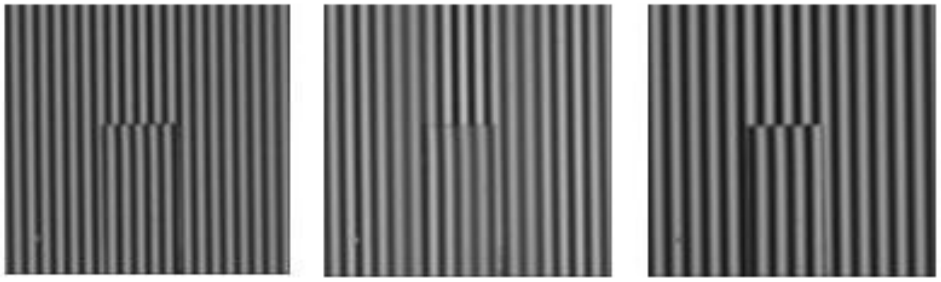


Fig. 8-3 Captured color image



(a) B (b) G (c) R

Fig. 8-4 Extracted B, G and R components

8.3 Crosstalk compensation method

8.3.1 Introduction

Since the R, G and B spectral sensitivity curves of the camera overlap each other, there is crosstalk between the signals. In particular, the crosstalk from the B-wave component to the G signal is estimated to be several tens of percent, so that the crosstalk compensation is indispensable to realize a three-wavelength single-shot measurement method using the G signal.

Crosstalk here refers to the fact that the R, G and B components of the camera do not completely correspond to the three illumination colors because the spectral sensitivity distributions of the R, G and B components of the color camera overlap, and, for example, the signal from the B illumination is mixed in with the G component of the camera.

An interference measurement method using a multi-wavelength simultaneous illumination source and a color camera such as the one in this paper is proposed by Pfortner et al. [2003]. However, the details of the optics have not been reported, and crosstalk compensation has not been implemented as unnecessary.

On the other hand, in the field of three-dimensional measurement by color projection, there are several reports on crosstalk compensation [Caspi et al. 1998, Huang et al. 1999, Sato et al. 2000, Zhang et al. 2006]. For example, Sato et al. [2000] performed crosstalk compensation between the RGB channels of an LCD projector and a color camera in their study of a rangefinder based on the color pattern projection method.

The author applied the compensation concept reported in the field of the above color projection method to single-shot interferometry, and obtained good results [Kitagawa 2009b]. We also studied how to check the effect of crosstalk compensation using the fact that the light sources of the interferometry are approximately monochromatic. In this section, we describe the algorithm of crosstalk compensation, how to obtain the crosstalk compensation coefficients, how to check the effect, and experimental results.

8.3.2 Crosstalk compensation method

As described in the previous section, we use the crosstalk compensation concept in the color projection method, and assume that crosstalk is represented by the following linear model:

$$\begin{cases} B' = B + aG + bR \\ G' = cB + G + dR \\ R' = eB + fG + R \end{cases} \tag{8–1}$$

where B', G' and R' are the observed intensities, B, G and R are the true intensities, and a, b, c, d, e and f are coefficients representing the crosstalk (called crosstalk coefficients, or coefficients). This model is derived from the fact that the RGB color model is an additive model and the assumption that the response of the camera is linear with the incident light intensity.

Since the coefficients are as small as a few percent except for c, if the product term of the coefficients is ignored, the true intensity can be calculated by the following equation,

$$\begin{cases} B = B' - aG' - bR' \\ G = -cB' + G' - dR' \\ R = -eB' - fG' + R' \end{cases} \tag{8–2}$$

8.3.3 Estimation of crosstalk coefficients

The following two methods were devised for estimating the coefficients[2)].

(1) Individual lighting method

Three color images obtained by turning on each LED light separately are shown in the first column of Fig. 8-5. When we decomposite these images, we can see the color crosstalk as shown in the second to fourth columns of Fig. 8-5. The correlation diagram shown in Fig. 8-6 is obtained from the R, B and G intensity values of each pixel of the color images. This figure shows the correlation between the B intensity and the G and R intensities when the B-LED is lit. From the regression coefficients, the crosstalk coefficients were obtained as c = 0.28 and e = 0.01. From similar experiments, the following coefficient matrix was obtained, which agrees with the values expected from the spectral characteristics in Fig. 8-2.

$$\begin{pmatrix} B' \\ G' \\ R' \end{pmatrix} = \begin{pmatrix} 1 & 0.05 & 0.00 \\ 0.28 & 1 & 0.04 \\ 0.01 & 0.04 & 1 \end{pmatrix} \begin{pmatrix} B \\ G \\ R \end{pmatrix} \tag{8–3}$$

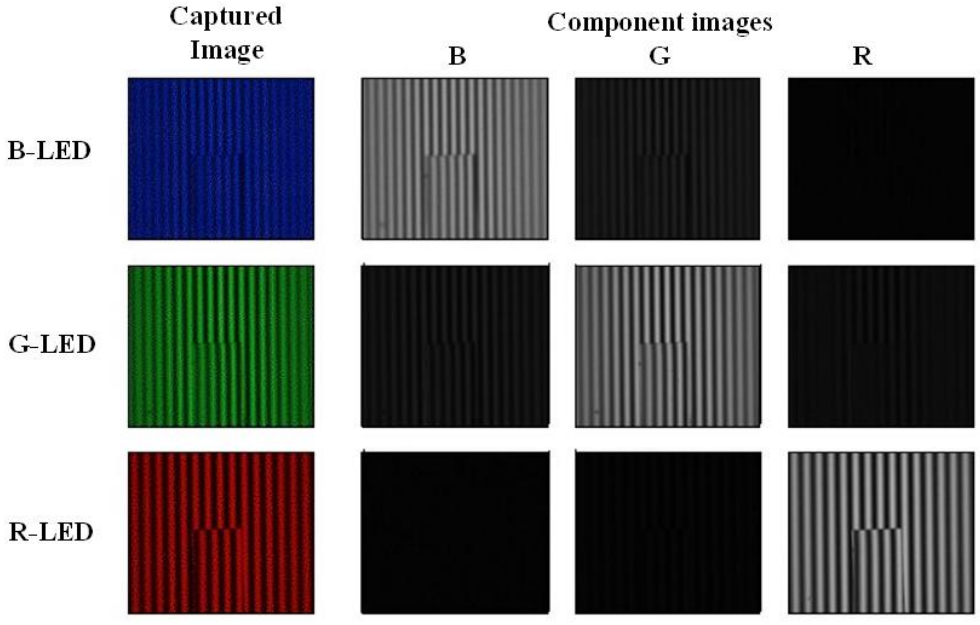


Fig. 8-5 Crosstalk effects between color channels

[2)] Estimating the crosstalk coefficients only needs to be performed once when the optical system is changed, and does not have to be performed for each measurement.

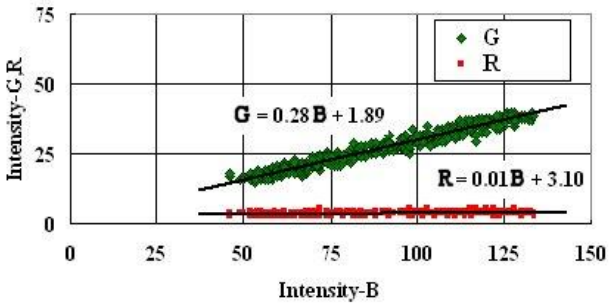


Fig. 8-6 Correlation among the RGB values before compensation

(2) All-lamps-lighting method

Crosstalk coefficients can be calculated from the images obtained by turning on all the lights. This method is simpler to operate than the individual lighting method, and has the feature that it can be applied even when individual lighting is not possible, such as in lighting systems using multi-wavelength bandpass filters described in Section 10.5.2.

In the three-wavelength single-shot interferometer, all three wavelength lights are turned on, and a sample with a homogeneous flat surface is selected as the measurement target. Then, since the DC component α, the interference amplitude γ, the phase ϕ and the frequency f (these are called the waveform parameters) of the interference fringes are assumed to be constant in the flat region, assuming the carrier fringes in the x direction, the intensities $B(x)$, $G(x)$ and $R(x)$ at a position x is represented by the following model:

$$\begin{cases} B(x) = \alpha_B + \gamma_B \cos(\phi_B + 2\pi f_B x) \\ G(x) = \alpha_G + \gamma_G \cos(\phi_G + 2\pi f_G x) \\ R(x) = \alpha_R + \gamma_R \cos(\phi_R + 2\pi f_R x) \end{cases} \tag{8–4}$$

On the other hand, the observed intensity affected by crosstalk is represented by the following crosstalk model derived from Eq. (8-1),

$$\begin{cases} B'(x) = B(x) + aG(x) + bR(x) \\ G'(x) = cB(x) + G(x) + dR(x) \\ R'(x) = eB(x) + fG(x) + R(x) \end{cases} \tag{8–5}$$

Substituting Eq. (8-4) into Eq. (8-5), the observed intensities at the coordinate x are expressed as a function of the waveform parameters and the crosstalk coefficients.

The crosstalk coefficients can be obtained by fitting the observed intensity model to the observed data by the least squares-method, i.e., by solving the least-squares problem with the following equation as the evaluation function,

$$J = \sum_{i=1}^{n} \left[\{B_i - B'(x_i)\}^2 + \{G_i - G'(x_i)\}^2 + \{R_i - R'(x_i)\}^2 \right] \tag{8–6}$$

where n is the number of data, and B_i, G_i and R_i (i = 1, n) are the observed intensities. There are a total of 18 unknown parameters: the waveform parameters for each wavelength, α_j, γ_j, ϕ_j and f_j (j = B, G, R) and the crosstalk coefficients a, b, …, f. In order to obtain these 18 parameters, we need more than 18 observed intensity data, which means that the number of

data n must be more than six.

Crosstalk coefficients were calculated using a single-shot image of a 1 μm standard step sample taken at three different wavelengths (Fig. 8-3). As the observed intensity values, we used the intensity values of ±100 pixels in the center of the image in x-coordinate on the y = 120 line of the image. The data are shown in Fig. 8-7. Solver in Excel was used for nonlinear least-squares fitting. The fitted intensity values are shown in Fig. 8-8. In addition, the obtained waveform parameters for each wavelength, α_j, γ_j, ϕ_j and f_j (j = B, G, R) are shown in Table 8-1, and the crosstalk coefficients a, …, f are shown in Eq. (8-7). These results are in good agreement with the results of the individual lighting method shown in Eq. 8-3, confirming the validity of the proposed method.

$$\begin{pmatrix} B' \\ G' \\ R' \end{pmatrix} = \begin{pmatrix} 1 & 0.05 & 0.00 \\ 0.30 & 1 & 0.04 \\ 0.00 & 0.06 & 1 \end{pmatrix} \begin{pmatrix} B \\ G \\ R \end{pmatrix} \tag{8–7}$$

Table 8-1 Obtained waveform parameters

	B	G	R
α	90	80	87
γ	69	71	74
ϕ	-0.15	-0.07	-0.09
f	0.0394	0.0344	0.0300

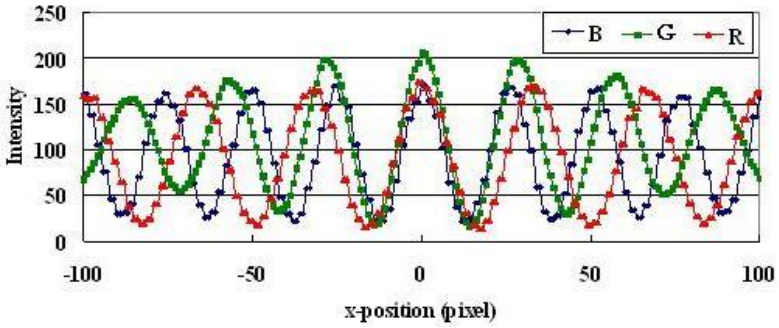


Fig. 8-7 Observed intensities

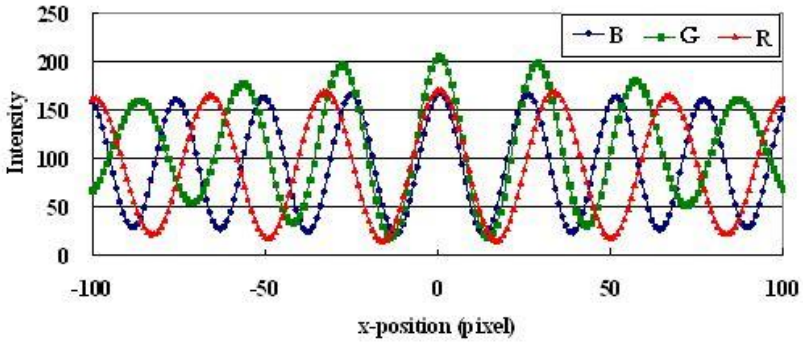


Fig. 8-8 Fitted values of intensities

8.3.4 Crosstalk compensation and its verification method

The crosstalk is compensated by using the coefficients obtained in the previous section by Eq. (8-2), and the effect of the compensation is confirmed by the following methods.

(1) Correlation of intensities

The correlations between the compensated B- and G-intensity values and R-intensity values are shown in Fig. 8-9. Compared with the uncompensated correlation diagram (Fig. 8-6), we can see that the crosstalk is completely removed.

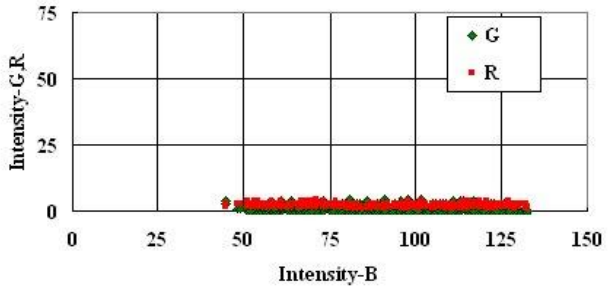


Fig. 8-9 Correlation among the RGB values after compensation

(2) Comparison of intensity profiles

The G intensity profiles before and after compensation are shown in Fig. 8-10. In the waveform before the compensation, the beat signal caused by the mixture of two frequencies is observed, but it disappears in the waveform after the compensation.

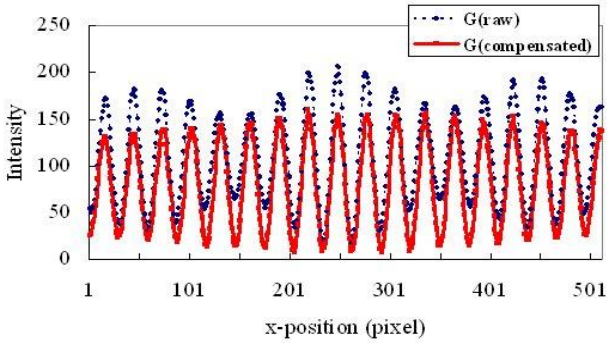


Fig. 8-10 Effect of crosstalk compensation: comparison of signal waveforms

(3) Frequency spectrum comparison

The frequency spectra before and after the compensation are shown in Fig. 8-11, and it was confirmed that the wavelength at around 470 nm was removed by the compensation.

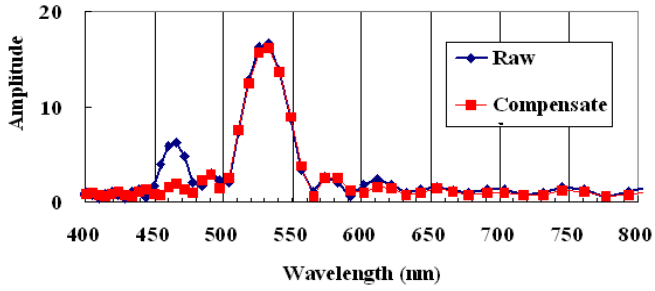


Fig. 8-11 Effect of crosstalk compensation: comparison of signal spectra

8.4 Frequency estimation method

8.4.1 Background

In order to realize the three-wavelength single-shot measurement method, high-precision frequency estimation is required. The frequencies are carrier fringe frequencies in the *x*- and *y*-directions, which are dependent on the relative tilt angle of the sample and reference planes and the wavelength.

The purpose of the frequency estimation is described below. The LMF method developed by the authors is used for the phase calculation of the interference fringe. The method uses the least-squares method to compute the three parameters (i.e. amplitude, phase and DC component) in the sinusoidal model, assuming that the frequency is known. Therefore, the frequency estimation is inevitable before the phase calculation.

Furthermore, the necessary accuracy of the frequency estimation is discussed. If there is an error in the frequency, the estimated phase changes linearly, as described in Section 8.4.4. When we consider the effect of this error on the measurement results, this is not a problem in the case of the single-wavelength method and the two-wavelength method, which uses the equivalent wavelength method for unwrapping, because the surface topography is measured only with an additional tilt. In the three-wavelength method, however, a small phase error may be magnified as a change in the fringe order because the coincidence method is used for unwrapping.

Many estimation methods have been proposed for the frequency of sine wave signals, such as the auto-correlation method, the Fourier transform method and the Prony method [Nadaya et al. 2000]. However, there is no practical method with high accuracy and low computational cost. Furthermore, when the fringes are parallel to the coordinate axis, the fringe frequency in the axial direction is close to zero, making it difficult to estimate with high accuracy using a one-dimensional estimation method.

The author devised a frequency estimation method using phase gradients and solved the above problem [Kitagawa 2009a]. In this section, we describe the principle of estimation, experimental results, implementation problems and solutions, and experimental results of actual samples.

8.4.2 Frequency estimation principle

(1) One-dimensional estimation

The proposed method (called the phase gradient method) uses the relationship between the frequency error and the phase gradient. Although this method can be applied to two-dimensional frequency estimation, the principle of this method is first described in one-dimensional estimation.

When we assume that the horizontal direction of the image with vertical fringes is the *x*-axis, the observed signal is represented by the following equation:

$$g(x) = A\cos(2\pi fx + \phi(x)) \tag{8–8}$$

where A is the amplitude, f is the frequency to be determined, and $\phi(x)$ is the phase at point x, which varies with the height.

The phase $\hat{\phi}(x)$ is obtained by fitting the observed signal to the following model function with the initial frequency estimate f_0.

$$\hat{g}(x) = \hat{A}\cos(2\pi f_0 x + \hat{\phi}(x)) \tag{8–9}$$

For this phase calculation, we use the LMF method described in Chapter 7.

On the other hand, Eqs (8-8) and (8-9) yield the following equation:

$$2\pi fx + \phi(x) = 2\pi f_0 x + \hat{\phi}(x) \tag{8–10}$$

Therefore, the phase $\hat{\phi}(x)$ is expressed as

$$\hat{\phi}(x) = 2\pi(f - f_0)x + \phi(x) \tag{8–11}$$

Considering a region that can be regarded as $\phi(x)$ = constant (called the reference plane region), the phase $\hat{\phi}(x)$ is linearly proportional to x, and its slope is proportional to the frequency error (f - f_0). Therefore, we can estimate the frequency from the following equation, which is obtained by differentiating both sides of Eq. (8-11)

$$\hat{f} = f_0 + \overline{(d\hat{\phi}(x)/dx)}/2\pi \tag{8–12}$$

Since the phase gradient $\overline{d\hat{\phi}(x)/dx}$ is obtained from at least two data points, the simplest way is to use the phase $\hat{\phi}_1$ and $\hat{\phi}_2$ at two different points x_1 and x_2 in the reference plane, and obtain the frequency $\hat{f}$ by the following equation;

$$\hat{f} = f_0 + (1/2\pi)(\hat{\phi}_2 - \hat{\phi}_1)/(x_2 - x_1) \tag{8–13}$$

The estimation accuracy can be further improved by using the obtained frequency as the initial value and repeating the estimation. In order to obtain the correct phase gradient, a phase unwrapping is usually required, and this problem is discussed in Section 8.4.7. For the derivation of Eq. (8-12), it was assumed that the reference region is a flat plane; however, it will be shown in Section 8.4.3 that the proposed method can be applied even in the absence of such a plane, i.e., when the reference region is not a flat plane.

(2) Two-dimensional estimation

Then, extending this method to two dimensions, Eqs (8-8), (8-9) and (8-11) are written as follows, respectively.

$$g(x, y) = A\cos(2\pi f_x x + 2\pi f_y y + \phi(x, y)) \tag{8–14}$$

$$\hat{g}(x, y) = \hat{A}\cos(2\pi f_{x0} x + 2\pi f_{y0} y + \hat{\phi}(x)) \tag{8–15}$$

$$\hat{\phi}(x, y) = 2\pi(f_x - f_{x0})x + 2\pi(f_y - f_{y_0})y + \phi(x, y) \tag{8–16}$$

The frequency estimation equation corresponding to Eq. (8-12) is as follows, which allows us to simultaneously estimate the frequencies in the x and y directions.

$$\begin{cases} \hat{f}_x = f_{x0} + (\overline{d\hat{\phi}(x,y)/dx})/2\pi \\ \hat{f}_y = f_{y0} + (\overline{d\hat{\phi}(x,y)/dy})/2\pi \end{cases} \tag{8–17}$$

8.4.3 Frequency estimation in non-flat areas

In this section, we show that the frequency estimation formula of Eq. (8-12), derived under the assumption that the frequency estimation region is flat, can be used even when the region is not flat. In this case, the apparent fringe frequency is not constant, but depends on the height variation. Therefore, we first calculate the estimation error due to the non-uniformity of the surface and then show that this does not affect the shape estimation.

From Eq. (8-11), the local phase gradient $d\hat{\phi}(x)/dx$ is given by

$$d\hat{\phi}(x)/dx = 2\pi(f - f_0) + d\phi(x)/dx \tag{8–18}$$

Therefore, from Eq. (8-12), the estimated frequency $\hat{f}$ is

$$\hat{f} = f + (\overline{d\phi(x)/dx})/2\pi \tag{8–19}$$

The second term on the right side of this equation indicates the error due to the non-flatness.

On the other hand, from the interference theory, the phase ϕ is

$$\phi(x) = 2\pi h(x)/(\lambda/2) \tag{8–20}$$

From Eqs (8-19) and (8-20), the estimated frequency $\hat{f}$ is expressed as

$$\hat{f} = f + (\overline{dh(x)/dx})/(\lambda/2) \tag{8–21}$$

This equation shows that the frequency error due to the non-flatness is proportional to the average gradient of the region and inversely proportional to the wavelength.

Next, the effect of this frequency error on the surface topography estimation is calculated. Letting $\hat{f}$ be the estimated frequency, Eq. (8-11) becomes

$$\hat{\phi}(x) = 2\pi(f - \hat{f})x + \phi(x) \tag{8–22}$$

On the other hand, the estimated height is given by

$$\hat{h}(x) = (\hat{\phi}(x)/2\pi)(\lambda/2) \tag{8–23}$$

Substitution of Eq. (8-22) into Eq. (8-23) yields

$$\hat{h}(x) = \{(f - \hat{f})x + \phi(x)/2\pi\}(\lambda/2) \tag{8–24}$$

From Eqs (8-21) and (8-24), we obtain

$$\hat{h}(x) = h(x) - (\overline{dh(x)/dx})x \tag{8–25}$$

This equation shows that the estimated height has a slope component and that the slope is proportional to the average slope of the region used for frequency estimation and that this is independent of the wavelength. Therefore, the non-flatness of the frequency estimation area does not affect the surface profiling. The slope component can be easily removed by fitting a

plane to the obtained height data.

8.4.4 Required accuracy for frequency estimation

This section discusses the required accuracy of the frequency estimation to correctly determine the fringe order when the coincidence method is used for three-wavelength unwrapping. For simplicity, we consider a 1-D measurement in the *x*-axis direction. We also focus on the absolute value of the error and ignore the positive and negative signs. Then the phase error $\Delta\phi$ due to the frequency error Δf is expressed by the following equation derived by a modification of Eq. (8-11).

$$\Delta\phi = 2\pi(\Delta f)x \tag{8–26}$$

The height error Δh due to phase error is defined as

$$\Delta h = \lambda(\Delta\phi)/4\pi \tag{8–27}$$

From Eqs (8-26) and (8-27), we obtain

$$\Delta h = (\lambda x/2)(\Delta f) \tag{8–28}$$

Since the maximum wavelength is about 600 nm and the maximum value of x is 256 pixels when the origin of the coordinate system is placed in the center of the image, the approximate maximum value of Δh is expressed as follows:

$$\Delta h \cong (300 \times 256)(\Delta f) \tag{8–29}$$

For phase unwrapping, we use the coincidence method, described later in Section 8.6. In this method, the height candidate values are obtained by changing the fringe order from the phase of each wavelength, and the combination in which the height candidate values of the three wavelengths best match is obtained.

Experience shows that the height error required to avoid incorrect order selection is about 10 nm. Therefore, if the allowable height error is 10 nm, the allowable frequency error is approximately 0.00013 (1/pixel) from Eq. (8-29). This is equivalent to 0.067 fringes in a 512-pixel image. Since the number of fringes in the experiments in this paper is about 15, the allowable relative error is about 0.4%.

8.4.5 Computer experiments (1): One-dimensional estimation

In order to verify the proposed method, computer experiments were conducted.

(1) Experimental conditions

The observed data were generated from the function $g(x) = \cos(2\pi fx)$. Here, we set the true frequency f = 0.020 and added the uniform noise of (− 0.1, + 0.1) to $g(x)$. The initial frequency f_0 = 0.019 (5% relative error), and the data size for the phase calculation was 25 pixels.

(2) Experimental results

Fig. 8-12 shows the observed data and the model function, where the amplitude A is set to 1 and the phase $\phi'(0)$ is set to 0. The phase $\phi'(x)$ obtained by the fitting and the regression equation are shown in Fig. 8-13. From the phase gradient of 0.00628, the frequency estimate is 0.019 + 0.00628/2 = 0.02000, which agrees with the true value with an error of less than 0.00001. Similar results were obtained for a wide range of initial frequencies.

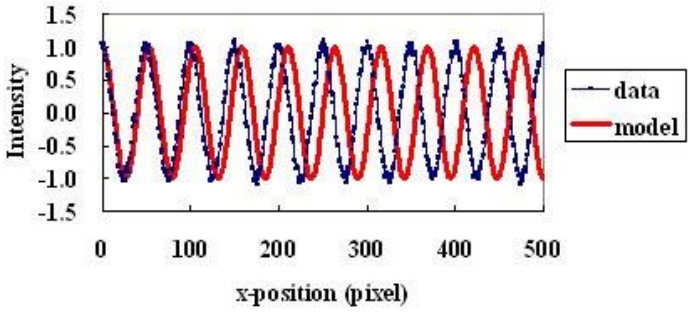


Fig. 8-12 Observed data and model function

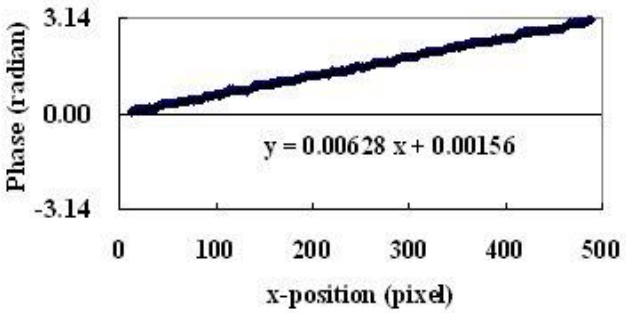


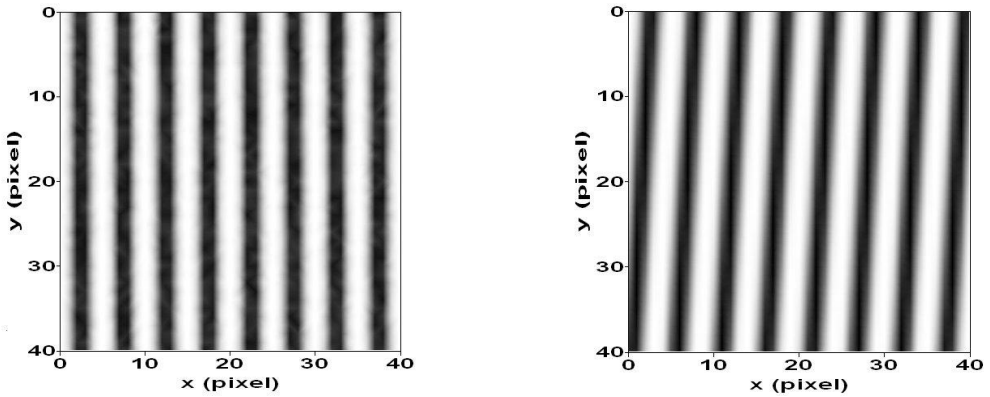
Fig. 8-13 Phase gradient

8.4.6 Computer experiments (Part 2): Two-dimensional estimation

Perform 2D frequency estimation for the case of zero frequency in the *y* direction. This is a case in which frequency estimation is extremely difficult with conventional 1D signal processing.

(1) Experimental conditions

The observed data were generated from the function $g(x, y) = \cos(2\pi f_x x + 2\pi f_y y) + 1$, where f_x = 0.2, f_y = 0.0, and the uniform noise of (- 0.1, + 0.1) is added to $g(x)$. The observed image is shown in Fig. 8-14(a). An error of 0.01 is expected for the initial frequency, and f_{x0} = 0.19, f_{y0} = 0.01. Figure 8-14(b) shows the model image when the amplitude A' = 1 and the phase $\phi'(x, y)$ = 0. The data for phase calculation is 3 × 3 pixels.



(a) Observed image (b) Model image with initial guess frequencies

Fig. 8-14 Images for two-dimensional frequency estimation

(2) Experimental results

The phase $\phi'(x, y)$ obtained by the fitting is shown in Fig. 8-15 and its profile is shown in Fig. 8-16(a)(b). The phase gradients in the x and y directions obtained by fitting the phase data to the plane $Z = aX + bY + c$ were 0.06275 and - 0.06298 (rad/pixel), respectively. From these data, we obtain the frequency estimates of f_x = 0.19999 and f_y = - 0.00002. The means ± standard deviations of ten repeated estimations were f_x = 0.20,000 ± 0.00003 and f_y = -0.00002 ± 0.00002. The same results were obtained in a wide range of initial frequencies. It is shown that a two-dimensional frequency estimation can be performed with high accuracy regardless of the direction and frequency of the fringes.

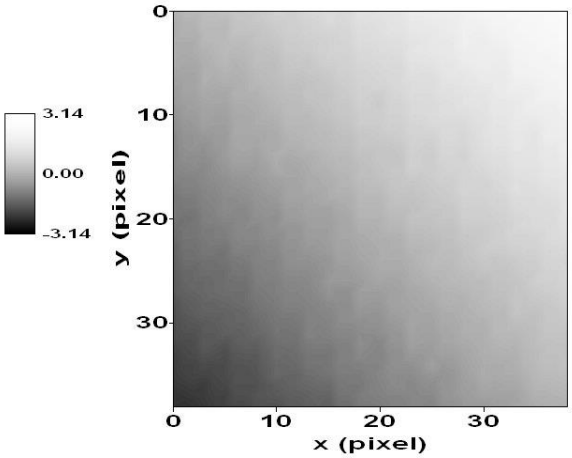
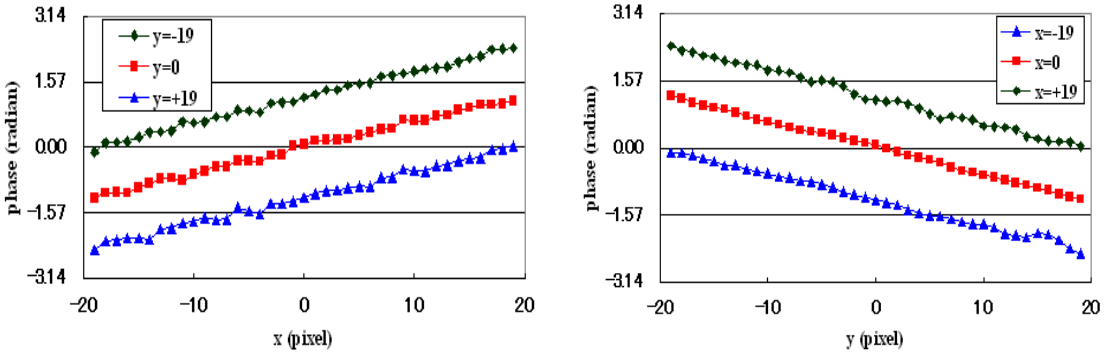


Fig. 8-15 Two-dimensional phase map



(a) *x*-direction (b) *y*-direction

Fig. 8-16 Phase profiles along three selected lines

8.4.7 Implementation problems and solutions

(1) Estimation flow and problems

The frequency estimation flow of this method is shown in Fig. 8-17. The dotted line shows a loop that repeats the estimation using the obtained estimates as the initial values. The condition of convergence is that the absolute values of the frequency compensation values f_x and f_y are less than or equal to the threshold for determining convergence. The threshold ε is set considering the fact that the required accuracy of frequency estimation is about 0.4% (see Section 8.4.4).

As mentioned in Section 8.4.2, phase unwrapping is usually required to obtain the correct phase gradient. The necessary example is shown in Fig. 8-18(a). Although the phase gradient is small, discontinuity is generated because the phase values are in the vicinity of ± π. There are many proposals for phase unwrapping, but robust phase unwrapping that are resistant to noise are computationally demanding. Therefore, we studied a method to make the phase unwrapping unnecessary.

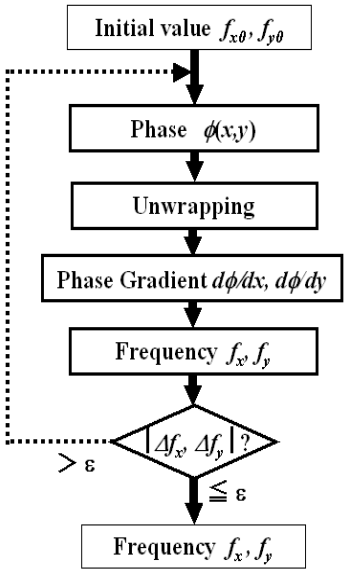
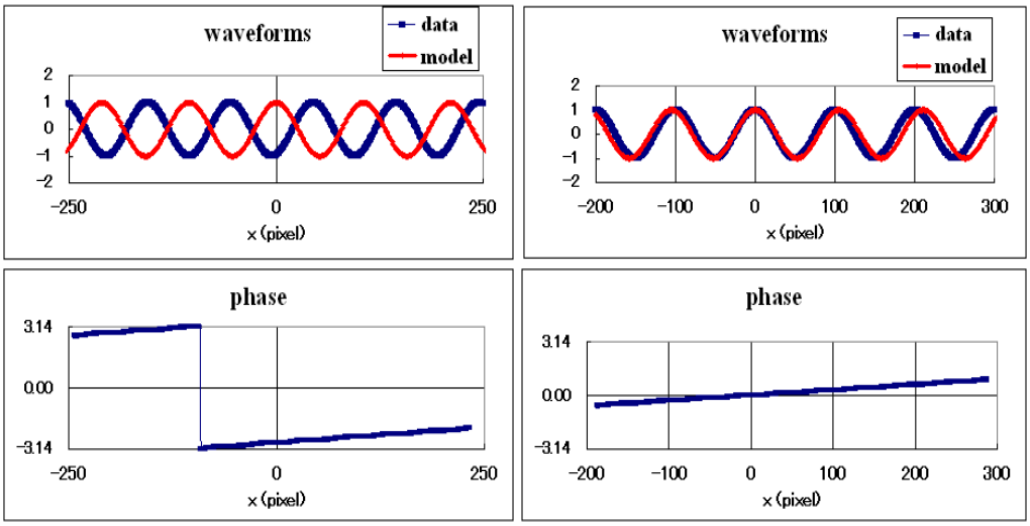


Fig. 8-17 Flowchart of frequency estimation

(2) Solution 1: Origin-shift method

In order to make the phase unwrapping unnecessary, it is effective to make the phase value obtained by the LMF method close to zero, which can be achieved by the method described below (called the "origin-shift method"). In other words, as shown in Fig. 8-18(b), when the phase of the model signal is matched to the phase of the observed signal, the calculated phase is around zero. In practice, the maximum location of the observed data signal is found, and the phase is calculated at the origin of the coordinate. This method eliminates the need for a phase unwrapping, as shown in Fig. 8-18(b).

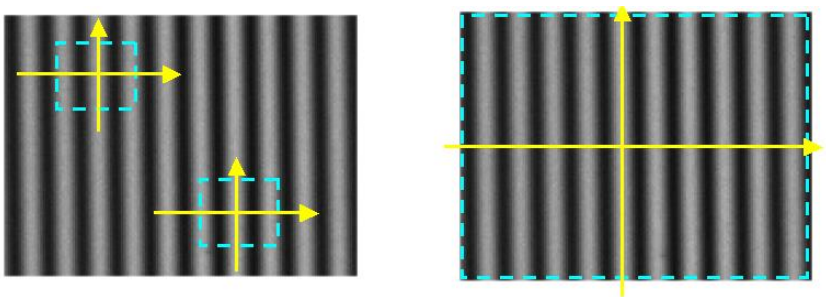


(a) With non-optimized origin (b) With optimized origin

Fig. 8-18 Effects of origin optimization

(3) Solution 2: Two-step method

A large initial error can lead to a large phase gradient, and a larger estimation region can lead to a phase exceeding $\pm\pi$. Therefore, as a first step, we perform a rough estimation in several small areas as shown by the blue dotted lines in Fig. 8-19(a). The solid yellow line in Fig. 8-19(a) shows the origin and the coordinate axes. Next, in the second step, we use the average of the frequencies obtained in the rough estimation as the initial value and perform the frequency estimation over a wide area, as shown by the dotted line in Fig. 8-19(b). Again, we adopt the "origin-shift method". This method eliminates the need for a phase unwrapping.



(a) 1st step: Rough estimation (b) 2nd step: Fine estimation

Fig. 8-19 Two-step estimation

8.4.8 Actual experiment

(1) Experimental method

The R component of the color images (Fig. 8-3) of the 1μm standard step sample is shown in Fig. 8-20. The rectangle in the lower center is a step.

(2) Experimental results

The rectangular area circled by the white line in the upper part of Fig. 8-20 was used as the reference plane area, and the frequency was estimated by varying the initial value. The data for phase calculation is 25 × 5 pixels. As a result, we obtained $f_x = 0.030064$, $f_y = 0.000361$ as the frequency estimates for the x and y directions with stable convergence over a wide range of initial values. This is 15.393 fringes in the x direction and 0.185 fringes in the y direction in terms of the number of fringes in a horizontally 512-pixel image.

Figure 8-21 shows the relationship between the initial value and the estimation results for single estimation (i.e. when the estimation is not repeated). The regression coefficient is 0.0029, which means that the frequency error is reduced to about 3/1000 by single estimation. When the convergence threshold is set at 0.4%, the estimation converges with a single estimation and does not need to be repeated.

The estimated time was 23 ms for a PC (Pentium 1.6 GHz). A sufficient accuracy was obtained even when the phase calculation area in the reference plane region was thinned out to 100 × 100 pixels instead of all pixels. The calculation time was reduced to 8 ms in this case.

8.4.9 Summary of frequency estimation

In this section, a phase gradient method is proposed as a new frequency estimation method for fringe images. The frequency is estimated by using the fact that the local phase value of the image depends on the frequency error. It is characterized by high accuracy, light computational cost, and simultaneous estimation of two-dimensional frequencies. It can also be applied to cases where the frequency is near zero. The validity and effectiveness of the proposed method were confirmed by computer and actual experiments. Furthermore, to solve implementation problems, some methods to eliminate the phase unwrapping were devised.

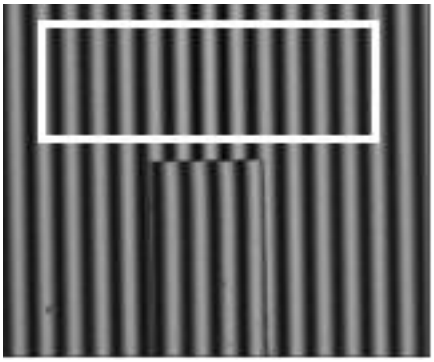


Fig. 8-20 Observed R image of 1μm step height
and its frequency estimation area

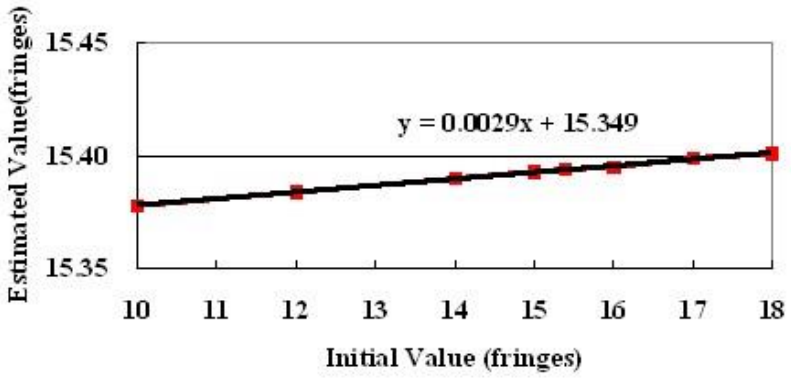


Fig. 8-21 Estimated frequencies versus initial values

8.5 Wavelength calibration method

8.5.1 Background

As shown in Table 8-1, the bandwidth of the light source used in this study is wide, ranging from 20 to 35 nm, and the center wavelength in the catalog is not exact. However, in order to extend the measurement range by the three-wavelength single-shot method, accurate wavelength values are required. This is because if there is an error in the wavelength, an error of the order multiplier will be introduced into the candidate height values, resulting in a wrong order estimation. Although the details are described in Section 8.5.3, the required wavelength accuracy is in the order of a few nm. Therefore, a method to calibrate each wavelength from the acquired image data was studied.

8.5.2 Algorithm

When the relative tilt angle of the reference and sample surfaces is defined as θ, the fringe frequency f is dependent on the light wavelength λ as follows:

$$f = (2\tan\theta)/\lambda \tag{8–30}$$

In other words, the wavelength of the light source is inversely related to the fringe frequency f. Since the fringe frequency can be obtained accurately by the method in Section 8.4, letting one wavelength be the reference wavelength λ_s, the other two wavelengths λ_i are calibrated relatively by the following equation

$$\lambda_i = (f_s / f_i)\lambda_s \tag{8–31}$$

This calibration should be performed only when the optical system is changed.

8.5.3 Wavelength accuracy considerations

In this section, we discuss the effect of wavelength error and the required accuracy. In the case of single-wavelength, the height is proportional to the wavelength, as shown in Eq. (6-14), and the height changes proportionally with wavelength error. This can be corrected by height calibration using a known sample.

In the case of two wavelengths, assuming that the equivalent wavelength method is used, if there is an error in the wavelength, the equivalent wavelength changes and then the estimated height changes proportionally. As in the case of the single-wavelength above, it can be corrected by height calibration.

In the case of three-wavelength method, the coincidence method described later in Section 8.6.2 is used to determine the order. In this case, the height error Δh due to the wavelength error can be calculated from the Eq. (6-14) as

$$\begin{aligned}\Delta h &= (\Delta\phi / 2\pi + n)(\Delta\lambda / 2) \\ &= (\Delta\phi / 2\pi)(\Delta\lambda / 2) + n(\Delta\lambda / 2)\end{aligned} \tag{8–32}$$

Since the second term is more dominant than the first term on the right side, it can be approximated as

$$\Delta h = n(\Delta\lambda / 2) \tag{8–33}$$

In the coincidence method, the height candidate values are obtained by changing the fringe order from the phase of each wavelength, and the combination in which the height candidate values of the three wavelengths best match is obtained. Experience shows that the height error required to avoid incorrect order selection is about 10 nm. Therefore, if the allowable height error is set to 10 nm, from Eq. (8-33), the allowable wavelength error is:

$$\Delta\lambda < 20 / n \tag{8–34}$$

In the experiments in this paper, the maximum height range is 4 μm and the maximum order n is about 15, so the allowable wavelength error is about 1.3 nm.

8.6 Three-wavelength phase unwrapping

8.6.1 Background

In the two-wavelength single-shot interferometry, the idea of "equivalent wavelengths" [Cheng et al. 1984] proposed for the two-wavelength phase shift method could be applied. However, in the case of three wavelengths, there is no established method for the phase shift method. The proposals so far can be divided into two categories:

(1) Based on the equivalent-wavelength method: There are three equivalent wavelengths when combined with two from three wavelengths. The data at the two wavelengths with the longest equivalent wavelengths are used to determine the order, and then the equivalent wavelength method is applied sequentially [Schneefuss et al. 2001, Warnasooriya et al. 2006, Meiners-Hagen et al. 2009]. In this case, the measurement range is estimated to be determined by the longest equivalent wavelength.

(2) Based on the coincidence method: The exact fraction method, which has been used for the calibration of block gauges [Tilford 1977, Yatagai 1988, Nishikawa et al. 1991, Iwata et al. 2003], is utilized.

Based on the comparative study of the two categories, this paper adopts the coincidence algorithm described below. This method can be applied not only to three wavelengths, but also to two or four wavelengths, and is highly versatile. The relationship between the coincidence method and the equivalent-wavelength method is discussed in detail in Section 9.2.

8.6.2 Algorithm of coincidence method

Letting i (i = 1, 2, 3) be the wavelength number and ϕ_i be the phase of wavelength λ_i, the candidate heights $h_i(n_i)$ are obtained by:

$$h_i(n_i) = (\phi_i / 2\pi + n_i)(\lambda_i / 2) \qquad (8\text{–}35)$$

where n_i is the integer order.

In the coincidence method, we find the combination of candidate height values with the lowest coincidence error within the expected height range, as shown in Fig. 8-22 [Tilford 1977, Yatagai 1988][3]. As an index of the coincidence error, the range (i.e. the difference between the maximum and minimum values), mean deviation, weighted mean deviation, variance, and standard deviation can be considered. However, based on the results of comparative experiments, it was judged that there is no significant difference between them, and we adopted the range expressed by the following equation as the "coincidence error" in this paper.

$$E(n_1, n_2, n_3) = \max\{h_1(n_1), h_2(n_2), h_3(n_3)\} - \min\{h_1(n_1), h_2(n_2), h_3(n_3)\} \qquad (8\text{–}36)$$

Furthermore, for the height calculation, we adopted the simple average of the three candidate height values expressed by

$$h = \{h_1(n_1) + h_2(n_2) + h_3(n_3)\} / 3 \qquad (8\text{–}37)$$

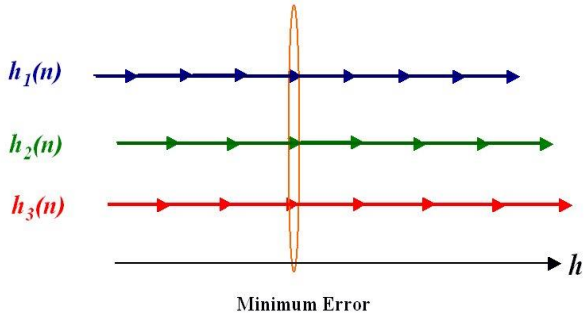


Fig. 8-22 Principle of exact fraction method

8.6.3 How to determine the order search range

In the above coincidence method, the height at which the coincidence error is minimized is sought within the expected height range. The search range is determined from the position of the zero-order fringe and the convexity range by the following procedure. Since the order is different for each wavelength, one of the wavelengths is used as the reference wavelength, and the order is expressed in terms of its order. In the following examples, the R wavelength is used as the reference wavelength.

(1) Detect the position of the zero-order fringe in the image.
(2) Obtain the order of the reference plane from the number of fringes between the position and the coordinate origin.
(3) Calculate the order range from the upper and lower height ranges.

An example is shown below, where the image is shown in Fig. 8-23, and the tilt is

[3] In the actual computation, we search for the minimum coincidence error by changing the combination of the orders n_i. Examples of specific calculations are shown in Figs. 9-4 and 9-11. A method to analytically determine this has also been studied. Tilford [1997] proposed a method to modify the estimate sequentially in length measurements using multi-wavelength laser interference. Agrok [1997, 2003] and Towers [2004a] have tried an interesting approach using the Chinese Remainder Theorem. However, these methods have many constraints and were not adopted in this paper.

downward to the right, and the estimated height range is − 4 ~ 0 μm of the reference surface.

First, the position of the zero-order fringes (x_0, y_0) in the image is detected with the center of the image as the origin. This can be achieved by detecting the position on the line perpendicular to the fringe direction (horizontal direction in this example) where the sum of the intensities of three wavelengths is at its maximum. Averages the positional coordinates over multiple lines, if necessary. In this example, $x_0 = 540$ is obtained. Next, the fringe order N is calculated by the following formula:

$$N = -(f_x x_0 + f_y y_0) \tag{8–38}$$

where f_x and f_y are the fringe frequencies of the reference wavelength, and the fringe order N corresponds to the number of fringes from the center (i.e. the origin) of the image to the zero-order fringe position. In this example, the frequencies of the reference wavelength are $f_x = -0.026$ and $f_y = 0$, from which $N = 14.0$ is obtained. This is made into an integer to obtain the reference surface order $n_0 = 14$.

Then, when the upper and lower height ranges from the reference plane are Δz_{up} and Δz_{down}, respectively, and the reference wavelength is λ_R, the order ranges n_{min} and n_{max} are represented by:

$$n_{min} = n_0 - \Delta z_{down}/(\lambda_R / 2)$$

$$n_{max} = n_0 + \Delta z_{up}/(\lambda_R / 2) \tag{8–39}$$

Here, each order must be an integer that includes a range of real values on the right-hand side. In this example, $\Delta z_{up} = 0$ nm, $\Delta z_{down} = 4000$ nm and $\lambda_R = 600$ nm yield $n_{min} = 0$ and $n_{max} = 14$.

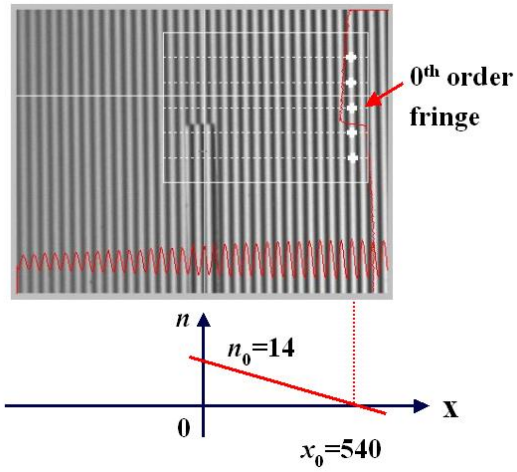


Fig. 8-23 Determination of fringe order

8.7 Experimental results and discussion

8.7.1 Experimental method

The illumination and imaging system of our surface profiler SP-500 (Fig. 2-10) was replaced with a three-wavelength imaging system, and the three-wavelength measurement software was installed in order to test the proposed method. A standard step sample (NTT-AT; HS-10) with a nominal step of 1120 nm was used as a sample for the test.

8.7.2 Fringe image capture

The images obtained by the proposed imaging system are shown in Fig. 8-3, and the separated BGR components are shown in Fig. 8-4. The intensity profiles of the y = 360 line are shown in Fig. 8-24.

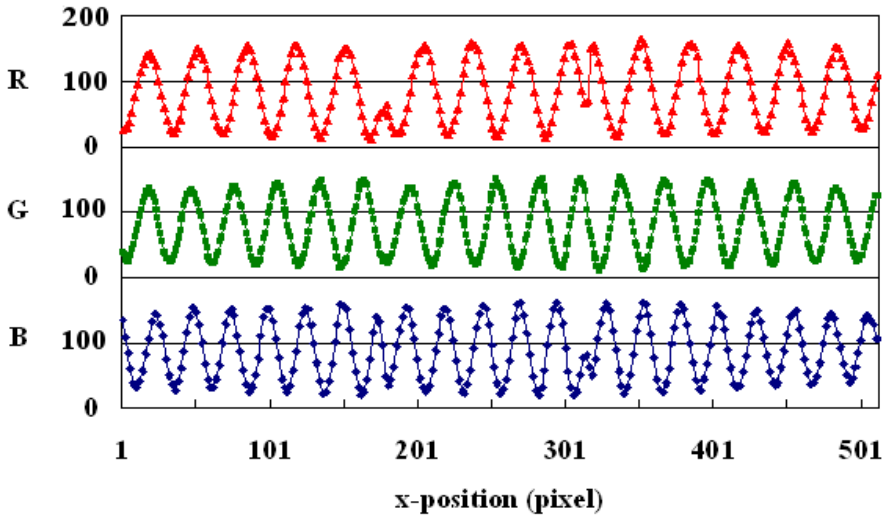


Fig. 8-24 Intensity profiles at y = 360

8.7.3 Crosstalk compensation

We compensated the crosstalk as described in Section 8.3. The crosstalk coefficients in Eq. (8-3) were used.

8.7.4 Frequency estimation

To estimate the carrier fringe frequency of each wavelength, the phase gradient method described in Section 8.4 was used. The reference area is a rectangle of 400 × 200 pixels on a flat surface at the top of the image. The phase calculation is based on the LMF method with a data size of 25 × 5 pixels. The results are shown in Table 8-2.

Table 8-2 Estimated frequencies

Color	f_x (1/pixel)	f_y (1/pixel)
B	0.0394	0.0005
G	0.0343	0.0005
R	0.0301	0.0004

8.7.5 Wavelength calibration

Using the results in the previous section and Eq. (8-31), we calibrated the other two wavelengths with the R wavelength as a reference. The results are shown in Table 8-3, which is in good agreement with the wavelength spectral peaks shown in Fig. 8-11.

Table 8-3 Results of wavelength calibration

Color	Calibrated value (nm)	Nominal value (nm)
B	478.5	470
G	548.9	530
R	-	627

8.7.6 Phase calculation

The phase of each wavelength was determined by the LMF method. The data size is 25 pixels wide by 5 pixels high. The phase profiles of each wavelength along the y = 360 line are shown in Fig. 8-25. The flat surface was measured without slope, which confirms the correctness of the set frequencies.

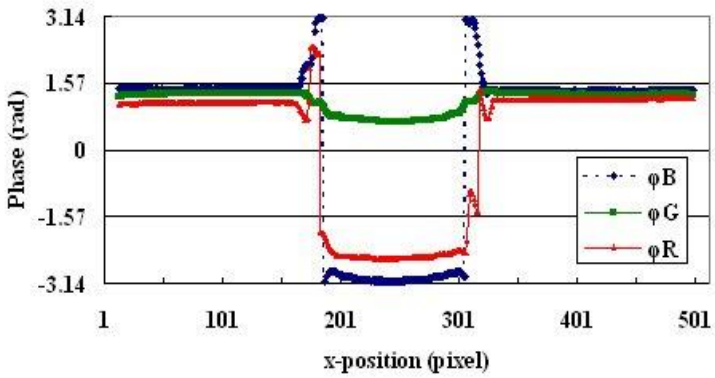


Fig. 8-25 Phase profiles at y = 360

8.7.7 Height calculation

The height profile of the y = 360 line is shown in Fig. 8-26 and the results of the 3D height calculation for the whole surface are shown in Fig. 8-27. A step of about 1100 nm has been correctly measured. The time required to calculate the height of the whole surface of 512×480 pixels was 0.42 s for a PC with an Intel 2.67 GHz processor.

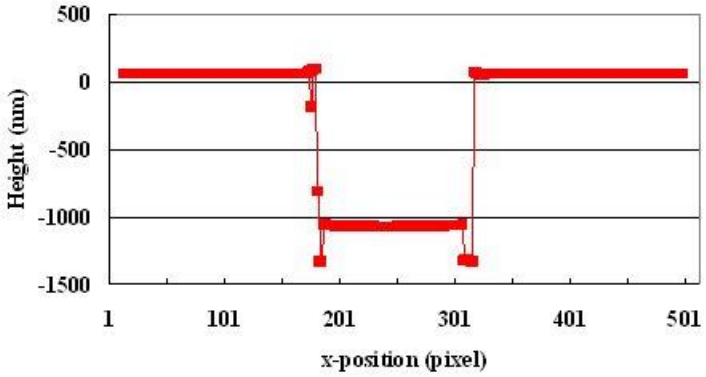


Fig. 8-26 Height profile at y = 360

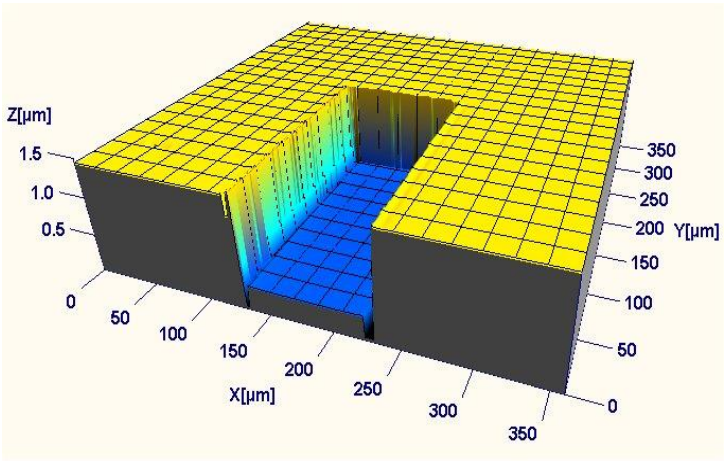


Fig. 8-27 Measurement result of 1 μm step height

8.7.8 Measurement reproducibility

We repeated the step measurements 10 times to determine the reproducibility. The step measurement method is based on the ISO 5436-14.[4)] As shown in Table 8-4, a reproducibility (σ) of 0.09 nm was obtained, confirming the high accuracy of this method.

Table 8-4 Results of step height repeatability test

No.	Step Height (nm)
1	1106.10
2	1106.04
3	1105.90
4	1105.83
5	1105.92
6	1105.88
7	1105.88
8	1105.80
9	1105.92
10	1105.88
Average	1105.10
σ	0.09

[4)] ISO 5436-1 specifies a method for determining a step from 2D data (line profile) obtained by a roughness tester or other means. In this paper, this specification is extended to three dimensions, and two planes parallel to the top and bottom data are fitted with a least-squares fitting, and the distance between the two planes is obtained as the step difference. The data size of each plane was 200 × 100 pixels.

8.8 Summary

In this chapter, a three-wavelength single-shot method is proposed as a new interference measurement method. An inexpensive and compact three-wavelength simultaneous imaging interference optical system is composed of a commercial LED lighting system and a color camera. From the color interference fringe image obtained by this system, the 3D shape is obtained by extracting the phase information for each wavelength. We have established the following techniques required:

(1) Crosstalk compensation method
(2) Frequency estimation method
(3) Wavelength calibration method
(4) Phase calculation method
(5) Three-wavelength unwrapping method

As a result of the above development, we were able to measure the surface topography of an actual sample with a steep step of more than 1000 nm.

This is an English translation of the article:
Katsuichi Kitagawa, "Chapter 8: Three-wavelength single-shot interferometry", in " Study on industrial surface profile measurement method based on optical interferometry", pp. 78-101, Doctoral Dissertation, University of Tokyo (2011)
(original in Japanese; translated and corrected by the author in Oct. 2020)