

Chapter 6: Local Model Fitting (LMF) Method

6.1 Research background

The Fourier transform method [Takeda et al. 1982] and the spatial phase synchronization method [Toyooka et al. 1984, Womack 1984, Kato et al. 1997]] are used to obtain the surface profile from a single fringe image in which the reference plane is inclined and carrier fringes are introduced. However, these fringe analysis methods are derived on the assumption that the frequency of the surface profile to be measured is sufficiently lower than the frequency of the carrier fringes, and thus have the problem that the high-frequency component becomes dull [Kato 1998]. To solve this problem, we developed a fringe image analysis algorithm named Local Model Fitting (LMF) [Sugiyama et al. 2006, Sugiyama et al. 2007].

6.2 Algorithm

The intensity of the fringe image corresponding to the point (x, y) on the sample surface is represented by $g(x, y)$. When the reference plane is tilted as shown in Fig. 1-9, $g(x, y)$ becomes

$$g(x, y) = a(x, y) + b(x, y) \cos(\phi(x, y) + 2\pi f_x x + 2\pi f_y y). \quad (6-1)$$

Here, $a(x, y)$ is the DC component of the interference fringe, $b(x, y)$ is the amplitude, and $\phi(x, y)$ is the phase. Also, f_x and f_y are carrier fringe frequencies in the x and y directions, respectively.

Since the phase $\phi(x, y)$ is a function of the surface height, the problem of finding the surface profile becomes the problem of finding $\phi(x, y)$.

In the LMF method, it is assumed that $a(x, y)$, $b(x, y)$ and $\phi(x, y)$ are constant near the point of interest (x_0, y_0) on the sample surface, and they are expressed as a , b , ϕ , respectively. Then Eq. (6-1) becomes:

$$g(x, y) = a + b \cos(\phi + 2\pi f_x x + 2\pi f_y y). \quad (6-2)$$

This is called the local model. If the surface to be measured is almost flat, there is no problem with the above assumption. On the other hand, this assumption does not hold in the steep step region, but its effect is limited to the neighboring region and does not propagate to other regions.

The fringe frequencies f_x and f_y are global variables that are determined by the relative tilt angle between the reference plane and the sample plane, and the wavelength and optical magnification. Therefore, it can be assumed to be constant in the image. Each estimated value is represented by f_x and f_y . These are pre-estimated by counting the number of stripes in the image.

As a result, the unknown variables in Eq. (6-2) are a , b and ϕ . In the LMF method, each estimated value \hat{a} , \hat{b} , $\hat{\phi}$ is determined by the least-squares method using intensity data of three or more points. That is, the intensity data $g(x_i, y_i)$ ($i = 1, 2, \dots, N$) ($N \geq 3$) are fitted to the local model of Eq. (6-2) to minimize the residual sum of squares expressed by:

$$J_1(a, b, \phi) = \sum_{i=1}^N [g(x_i, y_i) - a - b \cos(\phi + 2\pi \hat{f}_x x_i + 2\pi \hat{f}_y y_i)]^2. \quad (6-3)$$

The sample point (x_i, y_i) is selected from the vicinity of the point of interest (x_0, y_0) , but in the conventional fringe analysis method, it was general to take the sample point on a straight line. However, when using data with the same number of points, it is possible to localize the sample points more by distributing them in a two-dimensional manner than by taking them on a straight

line. In the LMF method, the sample point can be selected arbitrarily, so in the experiments in the next section, the sample point area is set to a circle or rectangle.

By the way, the model of Eq. (6-2) is nonlinear with respect to the variable ϕ . Therefore, the above fitting becomes a nonlinear least squares problem, and solving it as it is will increase the computational cost. Therefore, the following variable conversion is performed.

$$\xi_c = b \cos \phi, \quad (6-4)$$

$$\xi_s = b \sin \phi. \quad (6-5)$$

Also, to simplify the notation, we define:

$$\phi_c(x, y) = \cos(2\pi f_x x + 2\pi f_y y), \quad (6-6)$$

$$\phi_s(x, y) = -\sin(2\pi f_x x + 2\pi f_y y). \quad (6-7)$$

Then, Eq. (6-2) becomes a linear model as shown below:

$$g(x, y) = a + \xi_c \phi_c(x, y) + \xi_s \phi_s(x, y). \quad (6-8)$$

As a result, the problem of finding a, b, ϕ becomes a linear least squares problem of finding a, ξ_c, ξ_s , and the residual sum of squares is as follows:

$$J_2(a, \xi_c, \xi_s) = \sum_{i=1}^N [g(x_i, y_i) - a - \xi_c \phi_c(x_i, y_i) - \xi_s \phi_s(x_i, y_i)]^2. \quad (6-9)$$

The solution $\hat{a}, \hat{\xi}_c, \hat{\xi}_s$ to this problem can be obtained analytically by the following linear equation:

$$\mathbf{A} \begin{pmatrix} \hat{a} \\ \hat{\xi}_c \\ \hat{\xi}_s \end{pmatrix} = \mathbf{g}, \quad (6-10)$$

where

$$\mathbf{A} = \begin{pmatrix} N & \sum \varphi_c(x_i, y_i) & \sum \varphi_s(x_i, y_i) \\ \sum \varphi_c(x_i, y_i) & \sum \varphi_c(x_i, y_i)^2 & \sum \varphi_s(x_i, y_i) \varphi_c(x_i, y_i) \\ \sum \varphi_s(x_i, y_i) & \sum \varphi_s(x_i, y_i) \varphi_c(x_i, y_i) & \sum \varphi_s(x_i, y_i)^2 \end{pmatrix}, \quad (6-11)$$

$$\mathbf{g} = \begin{pmatrix} \sum g(x_i, y_i) \\ \sum g(x_i, y_i) \varphi_c(x_i, y_i) \\ \sum g(x_i, y_i) \varphi_s(x_i, y_i) \end{pmatrix}. \quad (6-12)$$

From $\hat{\xi}_c, \hat{\xi}_s$ obtained, the phase is calculated by the following equation:

$$\hat{\phi} = \text{atan}\left(\frac{\hat{\xi}_s}{\hat{\xi}_c}\right). \quad (6-13)$$

Here, the range of arctan is $-\pi/2$ to $\pi/2$, but it can be expanded to $-\pi$ to π by using the signs of the numerator and denominator in the arctan function.

The value obtained from Eq. (6-13) is the wrapped phase, and the height h is obtained by the following formula:

$$h = (\hat{\phi}/2\pi + n)(\lambda/2). \quad (6-14)$$

Here, n is an arbitrary integer, and some assumption is required for this determination. Generally, it is determined by phase unwrapping on the assumption that the phase difference between

adjacent pixels is π or less. That is, if the height difference between adjacent pixels is less than $\pm\lambda/4$, the correct height can be obtained from the phase.

The concept of the above LMF algorithm is shown in Fig. 6-1.

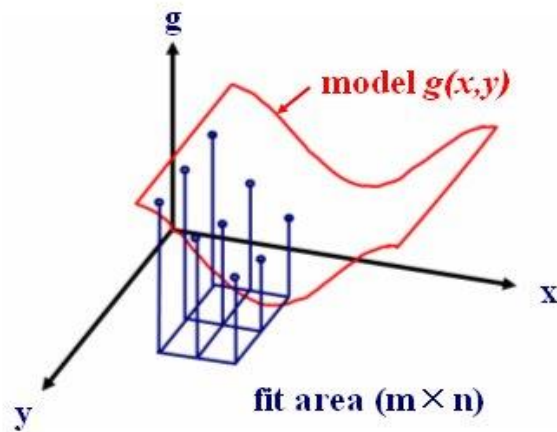


Fig. 6-1 Local Model Fitting method.

6.3 Implementation problems

We describe here some problems and solutions to implement the LMF algorithm on a computer.

6.3.1 Setting of the sample point area

The estimation accuracy of the LMF method depends on the selection of the sample point area used for fitting. If the area is too small, the amount of data used for parameter estimation will be small, making it susceptible to noise. On the other hand, if the area is too large, the estimation result will be too smooth, and the surface profile with steep steps cannot be measured correctly. In general, it is difficult to restore a profile that is too smooth to the original sharp profile, but it is possible to remove noise while maintaining the sharp profile. Therefore, as a standard, the sample point area is a rectangle of 25×5 pixels elongated in the carrier direction, and a median filter is applied to the obtained results to remove noise. Experimental results are shown in Sections 6.4 to 6.5.

6.3.2 Tilt direction and sign of frequency

When implementing the algorithm described in the previous section, the relative tilt direction of the sample plane and the reference plane, which causes carrier fringes, must be taken into consideration. For example, when an image with vertical stripes is given, the information about whether the sample surface is rising to the right (relative to the reference surface) or falling to the right (hereinafter simply called "tilt direction") is not included, so it is necessary to input this from outside.

We investigated a method to handle the tilt direction uniformly in the computer, and found a good way to associate the tilt direction with the positive/negative sign of the frequency. In Eq. (6-1), which is the basic expression of the interference fringe image, the coordinate axes x and y are assumed to be the x axis to the right and the y axis to the downward direction, as in normal image

expression. At this time, if the sign of the frequencies f_x (or f_y) is positive when the slope of the sample surface is positive along the x (or y) axis, and negative in the opposite case, then the height becomes proportional to the phase, and Eq. (6-14) can be used regardless of the tilt direction.

6.3.3 Effect of frequency estimation error

The proposed algorithm assumes that the fringe frequency is known. In the experiment, the number of stripes on the image was visually counted and input manually from the screen. The effect of this error is considered. For simplicity, consider the case where the carriers are introduced in the x direction. At this time, Eq. (6-2), which is the model equation of the LMF method, is

$$g(x) = a + b \cos(\varphi + 2\pi f x). \quad (6-15)$$

If there is a frequency error of Δf , then the frequency used for fitting is $\hat{f} = f + \Delta f$, and

$$g(x) = a + b \cos[\varphi - 2\pi(\Delta f)x + 2\pi(f + \Delta f)x]. \quad (6-16)$$

Then the obtained phase is

$$\hat{\varphi} = \varphi - 2\pi(\Delta f)x. \quad (6-17)$$

That is, a slope component proportional to the frequency estimation error Δf is mixed. However, this component does not pose a practical problem because it can be easily removed by fitting the plane to the final surface height and finding the difference.

6.3.4 Acceleration by sequential calculation

When calculating the phase at each point in the image, the LMF method is generally applied while moving the coordinates of the point of interest from the upper left to the lower right. If the algorithm in the previous section is used as it is, \mathbf{A} and \mathbf{g} will be calculated at each point using Eqs. (6-11) and (6-12). However, most of the elements of \mathbf{A} and \mathbf{g} are the sum of common terms at adjacent points. For example, when estimating using 125 points in a 5×25 rectangle (that is, $N = 125$), each element is the sum of 125 terms, but for two points that are horizontally adjacent, 120 terms are common.

Therefore, instead of newly calculating \mathbf{A} and \mathbf{g} at each point, the values already calculated at adjacent points can be used to reduce the amount of calculation. Figure 6-2 shows the case where the data area is 5 points vertical and 25 points horizontal. The leftmost 5 terms are subtracted from the sum of 125 terms at the left point, and the 5 rightmost terms are newly added.

In the proposed method, the first point is estimated by the conventional method. After that, it is repeated while moving in the horizontal direction, and when the point at the right end is estimated, it is shifted one line in the vertical direction.

The computational cost of this algorithm is compared with the conventional method. When the number of points used to estimate is $N = 5 \times 25$, the conventional method calculates 125 terms and adds 125 terms for each element. On the contrary, the new method needs 5 terms calculation and 10 terms addition/subtractions. Therefore, the computational cost of the proposed method is expected to be 1/10 or less that of the conventional method.

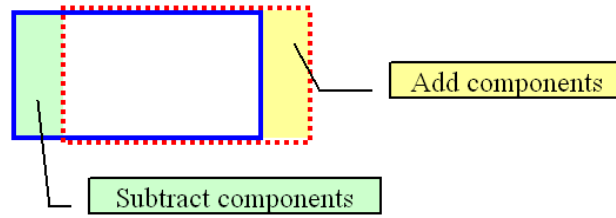


Fig. 6-2 Speed improvement of LMF calculation

6.4 Computer experiment

Assuming a step profile of 50 nm, we show that the original profile can be restored by the LMF method from the theoretical intensity data.

6.4.1 Experimental method

Assuming a one-dimensional surface profile $z(x)$ with a step of 50 nm (Fig. 6-3), the theoretical fringe value $g(x)$ is calculated by the interference fringe model equation $g(x) = \cos[2\pi f_x + 4\pi z(x)/\lambda]$. Here, the number of stripes is 10/512 pixels, and normal noise of $\sigma = 0.01$ is added to $g(x)$. Figure 6-4 shows the intensity profile. For this signal, the phase was calculated by changing the number of sample points N in the LMF method. Since the step difference is smaller than 1/4 of the wavelength, phase unwrapping is not performed, and the height is calculated with the fringe order $n = 0$ in Eq. (6-14).

6.4.2 Experimental results and discussion

Figure 6-5 shows the phase calculation results when the number of sample points N is 5. By converting this to height, Fig. 6-6 was obtained. When a 21-point median filter was applied to this, the results were as shown in Fig. 6-7, and the original profile was almost restored.

As described in Section 6.3.1, increasing the number of sample points N suppresses noise. The height measurement results for $N = 11$ and $N = 25$ (after the 21-point median filter) are shown in Figs. 6-8 and 6-9. There is a trade-off between noise suppression and horizontal resolution. In this experiment, the conditions shown in Fig. 6-8 seem optimal.

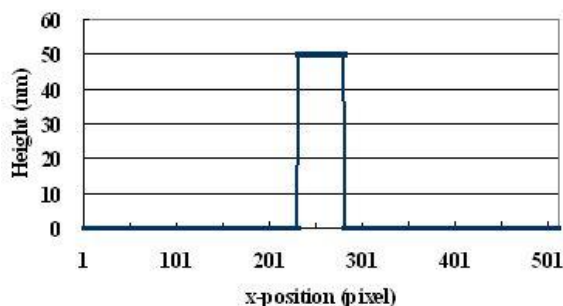


Fig. 6-3 Assumed height profile for computer experiments

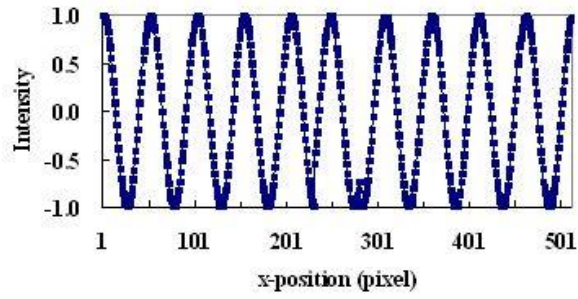


Fig. 6-4 Intensity profile

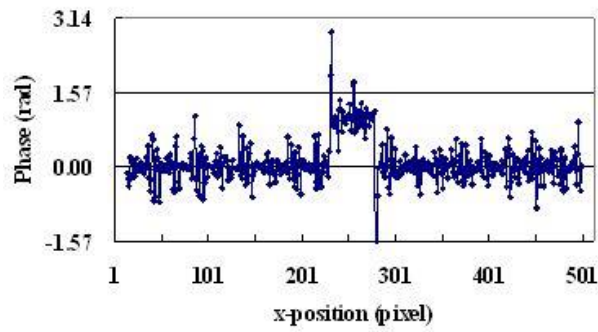


Fig. 6-5 Phase profile obtained with $N = 5$

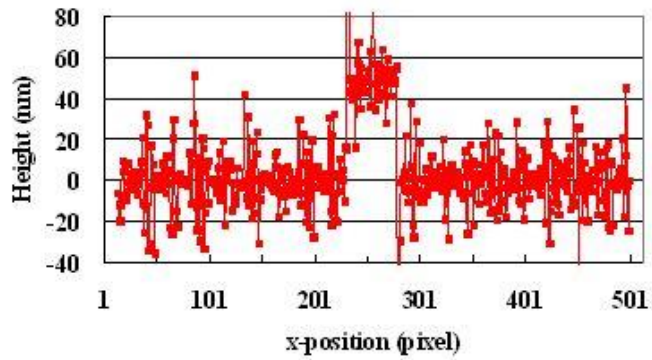


Fig. 6-6 Height profile obtained with $N = 5$

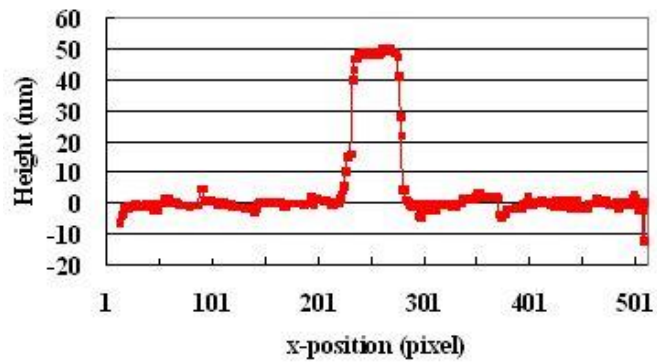


Fig. 6-7 Height profile with $N = 5$ after 21-point median filter

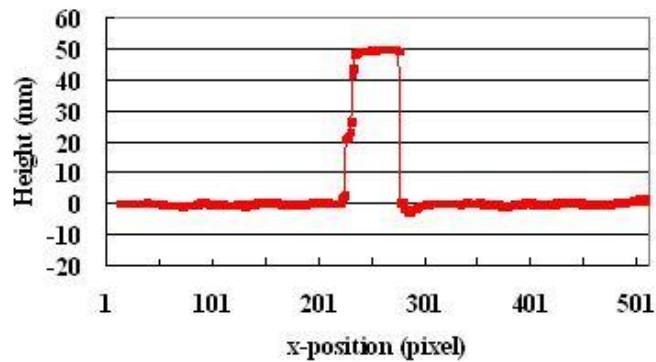


Fig. 6-8 Height profile with $N = 11$ after 21-point median filter

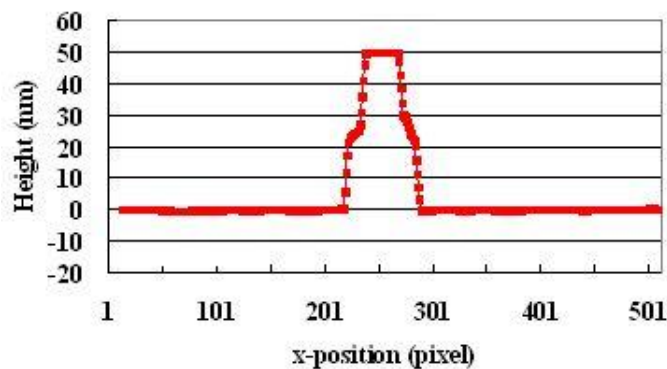


Fig. 6-9 Height profile with $N = 25$ after 21-point median filter

6.5 Actual experiments

In this section, the performance of the LMF method is evaluated using actual samples.

6.5.1 Experimental method

The surface profiler SP-500 [Kitagawa 2001a] shown in Figs. 2-10 to 2-11 was used for the experiment. The optical system was a Michelson interferometer, and an interference filter with a center wavelength of 600 nm and a bandwidth of 10 nm was inserted in the illumination unit. The measurement target is a standard step (made by NTT-AT) with a nominal value of 50 nm (manufacturer's actual measurement value is 53 nm) and has a steep step. The interference image was taken with a high-resolution (1344×1024 pixels) monochrome camera, and 512×480 pixels in the center of the image were extracted as the measurement target.

We implemented software for calculating height from captured image data on a personal computer (PC) using the LMF algorithm. The language used is C++. The sample point area for the LMF method is a rectangle centered on each point, and the vertical and horizontal sizes can be set freely. The phase unwrapping was based on the Minimum Spanning Tree method [Takeda et al. 1996] proposed for the phase shift method. This method extends the branches of the phase connection in the optimum direction considering the phase difference between adjacent pixels and the reliability of the phase, and has practical robustness.

6.5.2 Experimental results and discussion

The captured fringe image is shown in Fig. 6-10, and the intensity profile of the $y = 240$ line is shown in Fig. 6-11. As a measurement condition by the LMF method, a frequency that makes the result flat was searched and the number of stripes was set to 9.85. The data area used was 25 pixels wide \times 5 pixels high. Figure 6-12 shows the $y = 240$ line profile of the height measurement results, and Fig. 6-13 shows the full-image 3D display. The edge is measured as a steep profile. This result is in good agreement with the result of the conventional phase shift method (Fig. 6-14).

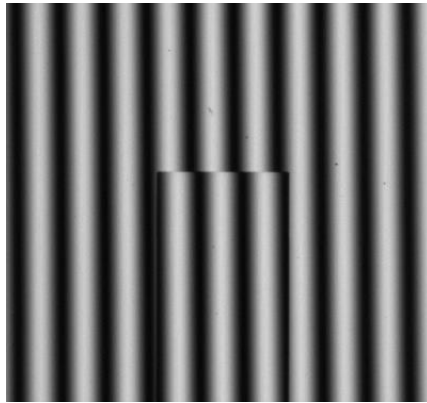


Fig. 6-10 Interferogram of 50nm step height

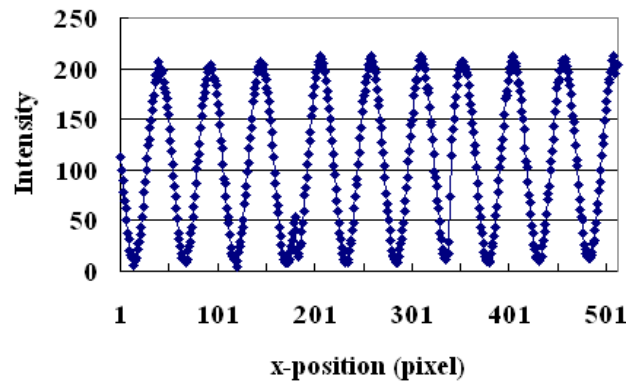


Fig. 6-11 Intensity profile at $y=240$

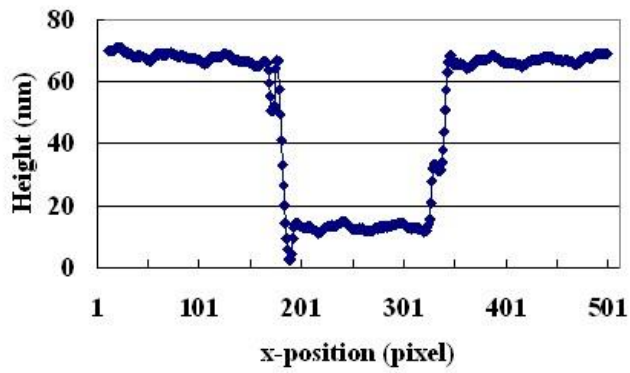


Fig. 6-12 Height profile at $y = 240$

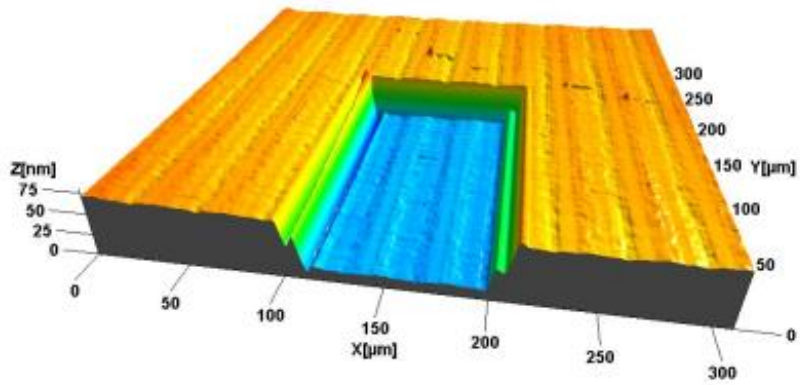


Fig. 6-13 3D profile of 50nm step height measured by LMF method

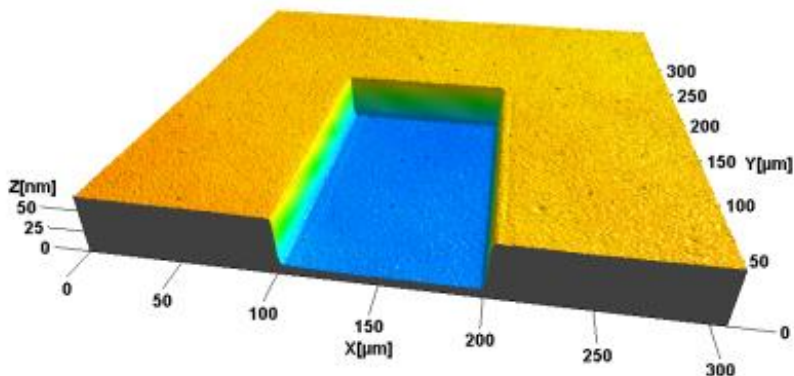


Fig. 6-14 3D profile of 50nm step height measured by PSI method

6.6 Comparison with conventional method

Next, the Fourier transform method [Takeda et al. 1982] was taken up as a conventional method, and a comparative experiment with the proposed method was performed. The principle of the Fourier transform method is shown in Fig. 6-15. When the intensity profile shown in Fig. 6-11 is

Fourier transformed, the frequency spectrum shown in Fig. 6-16 is obtained. The phase profile is obtained by extracting the carrier frequency component, shifting it to frequency zero, and then performing an inverse Fourier transform. Converting this to height, Fig. 6-17 was obtained. As expected from the principle of the Fourier transform method, a steep step is blunt because only a frequency component much lower than the carrier frequency is reproduced.

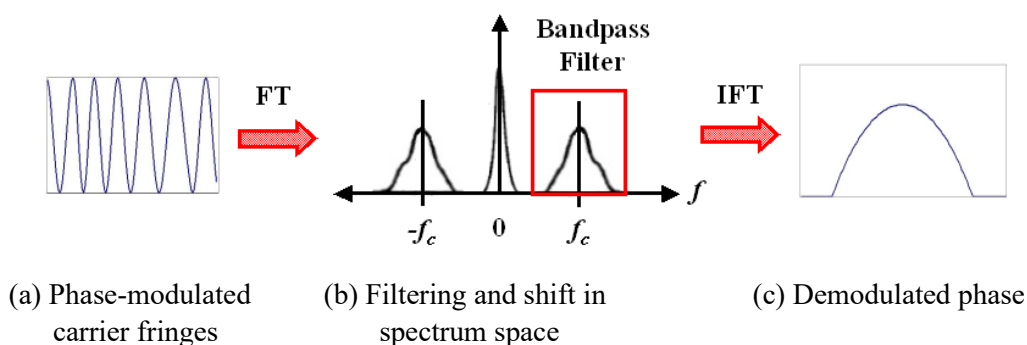


Fig. 6-15 Fourier-transform method

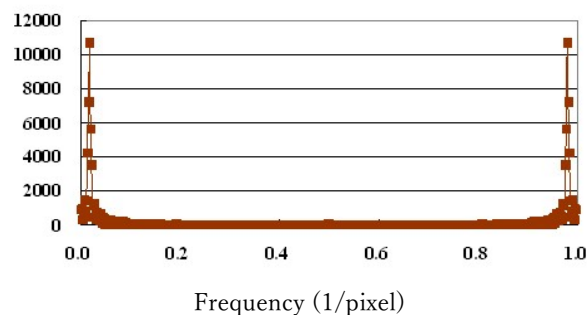


Fig. 6-16 Fourier spectrum of 50 nm step height

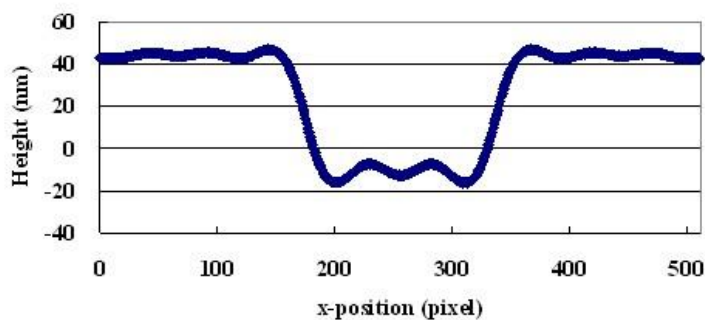


Fig. 6-17 Height profile of 50 nm step height obtained by Fourier-transform method

6.7 Improvements of LMF method

The algorithm described in Section 6.2 is the basic algorithm of the LMF method. After 2006, the following improved algorithms have been studied and proposed in order to reduce the disturbance near the step and to speed up the calculation. However, since it is not used in the

multiwavelength one-shot method, which is the subject of this paper, it will not be described in detail.

- (i) LMF with interpolation (iLMF): Fits at points with small error and interpolates at other points [Yokota et al. 2009].
- (ii) Weighted LMF (wLMF): The weight of the least squares method is changed according to the distance from the target point.
- (iii) Iterative reweighting LMF method (IRLMF): The weights of the least squares method are adaptively determined according to the estimated height of the region near the point of interest [Kurihara et al. 2010, Kurihara et al. 2010].

6.8 Summary

In this chapter, we proposed a local model fitting method (LMF method) as a new fringe image analysis algorithm. This method is characterized by less high-frequency component blunting and lower computational cost than the conventional Fourier transform method. Through computer experiments and actual sample experiments, it was confirmed that the surface profile with a sharp step difference of 50 nm can be measured correctly.

This is an English translation of the article:

Katsuichi Kitagawa, "Chapter 6: Local Model Fitting (LMF) Method",
in "Study on industrial surface profile measurement method based on optical
interferometry", pp. 61-71, Doctoral Dissertation, University of Tokyo (2011)
(original in Japanese; translated and corrected by the author in Aug. 2020)