This document is downloaded from DR-NTU (https://dr.ntu.edu.sg)
Nanyang Technological University, Singapore.

Improvement of least squares integration method with iterative compensation for shape reconstruction from gradient

Huang, Lei; Asundi, Anand Krishna

2012

Huang, L., & Asundi, A. K. (2012). Improvement of least squares integration method with iterative compensation for shape reconstruction from gradient. Proceedings of SPIE - Optical Micro- and Nanometrology IV, 84300S.

https://hdl.handle.net/10356/98855

https://doi.org/10.1117/12.921402

© 2012 SPIE. This paper was published in Proceedings of SPIE - Optical Micro- and Nanometrology IV and is made available as an electronic reprint (preprint) with permission of SPIE. The paper can be found at the following official DOI: [http://dx.doi.org/10.1117/12.921402]. One print or electronic copy may be made for personal use only. Systematic or multiple reproduction, distribution to multiple locations via electronic or other means, duplication of any material in this paper for a fee or for commercial purposes, or modification of the content of the paper is prohibited and is subject to penalties under law.

Downloaded on 13 Mar 2024 18:52:09 SGT

Improvement of least squares integration method with iterative compensation for shape reconstruction from gradient

Lei Huang, Anand Asundi

School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore 639798, Singapore

Abstract: An improvement is made to the traditional 2D integration with least squares method by introducing an iterative compensation procedure. The issue of inaccurate reconstruction due to imperfection of Southwell grid model is solved through the introduced iterative compensations. The feasibility and superiority of the proposed method are investigated with simulations. Moreover, the proposed method is compared with the integration method with radial basis functions.

Keywords: Least squares method, iterative compensation, two-dimensional integration, shape from gradient

1. Introduction

A variety of optical techniques are explored in three-dimensional (3D) shape metrology for industrial inspection applications [1]. Some of these shape measurement techniques are referred to direct methods, since they get 3D results directly from the captured optical signals, *e.g.* laser spots or pattern images. Other techniques deduce the shape by integration of the surface gradient information. The fringe reflection technique [2-3] and Shack-Hartmann methods [4] are two representatives of these indirect shape measurement methods. Consequently, a process to integrate the shape from gradient data is required in order to reconstruct the shape of the surface or wave front.

The shape or profile results are influenced by the two-dimensional (2D) integration process, and hence research efforts to make this process more accurate and faster are being explored. Fried [5], and Hudgin [6-7] studied the reconstruction of wave front from phase difference or slope using a least squares fitting approach, and evaluated the error propagation in reconstructions. Southwell [8] provided the wave front estimation from slope data with least squares method in more details including a discussion on grid models, solutions, and errors of least squares integration method. Li *et al.* [9] reviewed several evaluation methods for gradient measuring techniques and a comparison of these integration methods shows the superiority of the least squares integration method in terms of accuracy, robustness and flexibility. Ettl *et al.* [10] proposed another accurate shape reconstruction method from gradient data by dividing the whole area into in small pieces with certain common regions, integrating the slope data piece-wise by radial basis functions (RBFs) with least squares sense, and then stitching all pieces together to reconstruct the entire surface.

In least squares integration method, the widely-used Southwell grid model assumes the surface can be expressed as a biquadratic spline function. This assumption is obviously not satisfied in many situations resulting in integration errors. Ettl's RBFs based method is not able to directly handle large matrices without dividing and stitching process and it is slow even for one small piece of data with RBFs, although it may perform with high accuracy and flexibility.

Least squares integration method with Southwell grid model [8] is basically good for handling large datasets, because sparse matrices can be utilized to reduce the memory usage during the least squares integration. However, the accuracy of least squares integration method is affected by its assumption which is not commonly satisfied. Consequently, this work aims to improve the integration accuracy due to the imperfection of the assumption from Southwell grid model.

The rest of this article is organized as follows. Section 2 briefly recalls the least squares integration with Southwell model at first, and then describes the proposed iterative compensation process to increase the integration accuracy. Section 3 demonstrates serious of simulations to show the feasibility of the novel iterative least squares integration method. Section 4 discusses merits and limitations of the new method with a performance comparison of RBFs based method. Section 5 concludes the work.

Optical Micro- and Nanometrology IV, edited by Christophe Gorecki, Anand K. Asundi, Wolfgang Osten, Proc. of SPIE Vol. 8430, 84300S · © 2012 SPIE · CCC code: 0277-786X/12/$18 · doi: 10.1117/12.921402

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 02/05/2014 Terms of Use: http://spiedl.org/terms

2. Principle

In some slope-measuring-based optical metrology methods, integration converts gradient data to surface shape using least squares fitting. Following Southwell grid model [8] as shown in Fig. 1, the relationship between slope and shape can be expressed as

$$\begin{cases} \dfrac{p_{i,j+1}+p_{i,j}}{2}=\dfrac{z_{i,j+1}-z_{i,j}}{x_{i,j+1}-x_{i,j}}, & i=1,\dots,M, \quad j=1,\dots,N-1 \\ \dfrac{q_{i+1,j}+q_{i,j}}{2}=\dfrac{z_{i+1,j}-z_{i,j}}{y_{i+1,j}-y_{i,j}}, & i=1,\dots,M-1, \quad j=1,\dots,N \end{cases}, \tag{1}$$

where x, y, z are the coordinates. p and q are the slopes in x- and y-directions, respectively, *i.e.* $p = dz/dx$ and $q = dz/dy$.

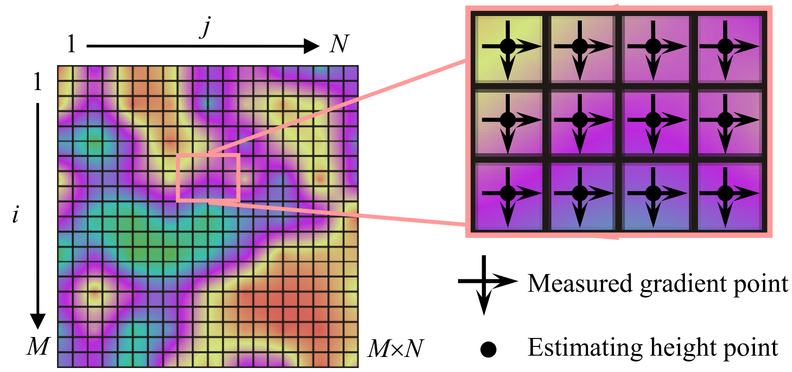


Fig. 1. The Southwell grid model in least squares integration.

Therefore the height information can be given as

$$\begin{cases} z_{i,j+1}-z_{i,j}=\dfrac{1}{2}\left(p_{i,j+1}+p_{i,j}\right)\left(x_{i,j+1}-x_{i,j}\right), & i=1,\dots,M, \quad j=1,\dots,N-1 \\ z_{i+1,j}-z_{i,j}=\dfrac{1}{2}\left(q_{i+1,j}+q_{i,j}\right)\left(y_{i+1,j}-y_{i,j}\right), & i=1,\dots,M-1, \quad j=1,\dots,N \end{cases}. \tag{2}$$

In terms of matrices, it shows as

$$\mathbf{DZ}=\mathbf{G}, \tag{3}$$

where these matrices are

$$\mathbf{D}=\left[\begin{array}{ccccccccc} -1 & 0 & \cdots & 0 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & -1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & 0 & -1 & 0 & \cdots & 0 & 1 \\ \hline -1 & 1 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & -1 & 1 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & -1 & 1 \end{array}\right], \tag{4}$$

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 02/05/2014 Terms of Use: http://spiedl.org/terms

$$\mathbf{Z}=\begin{bmatrix} z_{1,1} \\ z_{2,1} \\ \vdots \\ z_{M,N} \end{bmatrix}, \tag{5}$$

$$\mathbf{G}=\frac{1}{2}\begin{bmatrix} \left(p_{1,2}+p_{1,1}\right)\left(x_{1,2}-x_{1,1}\right) \\ \left(p_{1,3}+p_{1,2}\right)\left(x_{1,3}-x_{1,2}\right) \\ \vdots \\ \left(p_{M,N}+p_{M,N-1}\right)\left(x_{M,N}-x_{M,N-1}\right) \\ \hline \left(q_{2,1}+q_{1,1}\right)\left(y_{2,1}-y_{1,1}\right) \\ \left(q_{3,1}+q_{2,1}\right)\left(y_{3,1}-y_{2,1}\right) \\ \vdots \\ \left(q_{M,N}+q_{M-1,N}\right)\left(y_{M,N}-y_{M-1,N}\right) \end{bmatrix}. \tag{6}$$

Commonly Eq. (3) is over determined in practical situations. Hence, a least squares estimation can be carried out to provide the height distribution as an optimized solution.

The assumption implied in Eq. (1) is that the slope distribution is bilinear within each tiny quadrilateral, *i.e.* the variation of surface height is biquadratic. As we mentioned above, this assumption is not always satisfied, and hence is a main source of error. An iterative compensation procedure is thus proposed to improve the integration accuracy. Basically higher order terms are contained in these residual gradient data, which cannot be fit with the biquadratic spline. By integrating these residual gradient data it is possible to compensate errors due to the higher order terms. Following a few iterations, the final shape can be accurately reconstructed with no influence of the imperfect assumption.

The specific steps of the algorithm are described as follows.

Step 1: Integrate the gradient data with the traditional least squares integration method to get initial height $z_0(x, y)$. Set the current height $z(x, y) = z_0(x, y)$.

Step 2: Calculate the residual slopes in x- and y-directions $dp(x, y)$ and $dq(x, y)$ with current height distribution $z(x, y)$.

Step 3: Integrate the residual slopes $dp(x, y)$ and $dq(x, y)$ with the traditional least squares integration method to get $z_c(x, y)$ for compensation.

Step 4: Compensate the height $z(x, y) = z(x, y) + z_c(x, y)/n_k$, where the parameter n_k are empirical constants which can be determined through simulation and k stands for the number of compensation.

Step 5: Repeat the compensation loop from step 2 to step 4 till the compensating term $z_c(x, y)$ is less than the preset threshold z_{thr}. The current $z(x, y)$ is the final height.

Note: the empirical parameters n_k are determined by simulations with surfaces containing different higher order components. Form the numerical calculation, these empirical parameters n_k are $n_1 \approx 3$, $n_2 \approx 4.0909$, $n_3 \approx 4.9476$, $n_4 \approx 5.6768$ when k=1, 2, 3, 4.

3. Simulation

A series of simulation were conducted in order to investigate the validity of the proposed iterative compensation. First, an example is to show the iterative compensation works efficiently to improve the integration accuracy when the assumption of biquadratic spline is not satisfied. Both coordinates of x and y range from -5mm to 5 mm with an interval

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 02/05/2014 Terms of Use: http://spiedl.org/terms

of 0.02mm, *i.e.* the size of matrices is 500 by 500, which yields 500,000 x- and y-slope values in all. As shown in Fig. 2(a), the true out-of-plane dimension is predefined as

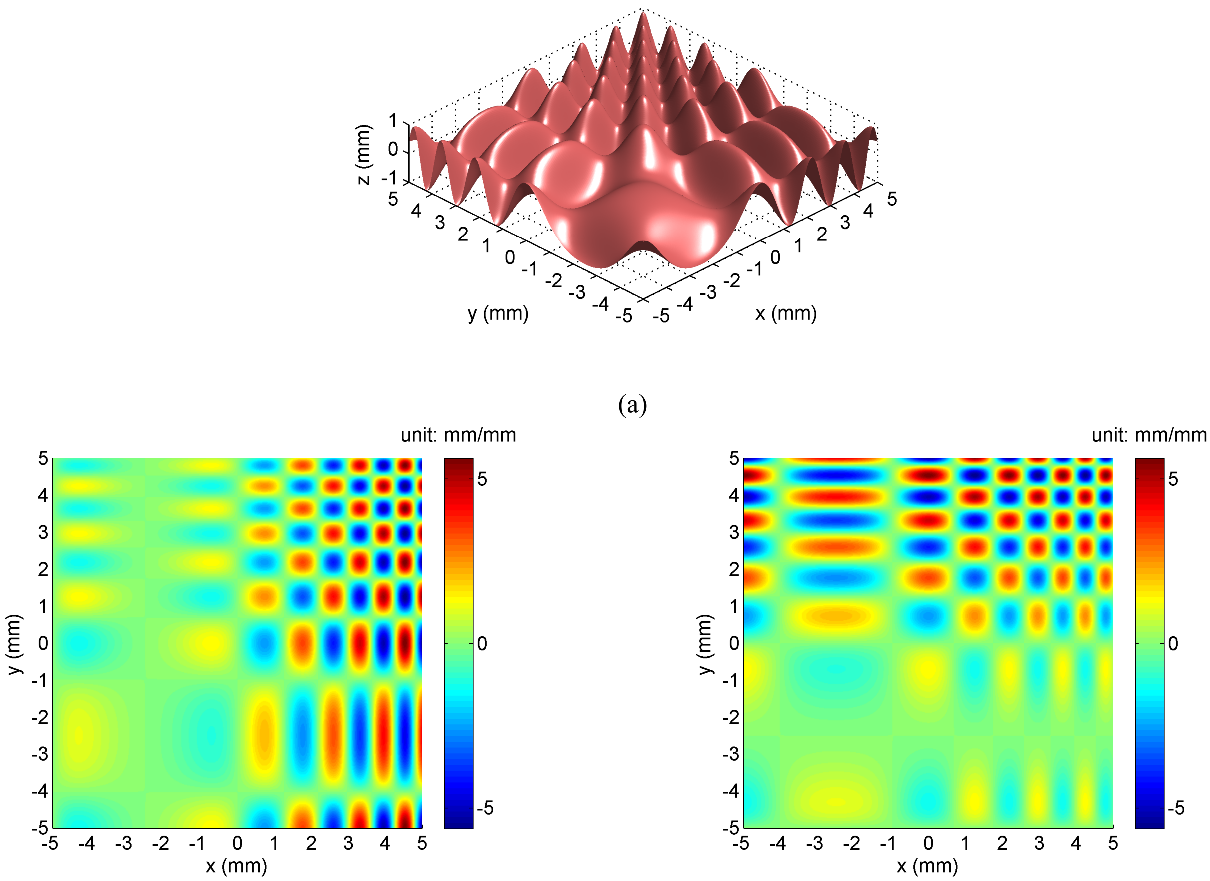
$$z = \cos\left(0.4x^2 + 2x\right)\cos\left(0.4y^2 + 2y\right), \tag{7}$$

and its corresponding slopes in x- and y-directions p and q shown in Fig. 2(b) and (c) can be derived as

$$p = -\left(0.8x + 2\right)\sin\left(0.4x^2 + 2x\right)\cos\left(0.4y^2 + 2y\right), \tag{8}$$

$$q = -\left(0.8y + 2\right)\cos\left(0.4x^2 + 2x\right)\sin\left(0.4y^2 + 2y\right). \tag{9}$$

In Fig. 2(b) and (c), the surface gradient varies slightly at the lower left corner and acutely at the upper right corner. The sample surface shape is specially chosen by purpose to test the iterative performance of the proposed method.



(a)

(b) (c)

Fig. 2. True shape and gradient data in simulation. (a) True shape, and true slopes in x-direction $p = \mathrm{d}z/\mathrm{d}x$ (b) and in y-direction $q = \mathrm{d}z/\mathrm{d}y$ (c).

By applying the proposed integration method with iterative compensation, the out-of-plane height distribution can be accurately reconstructed as shown in Fig. 3. The result from traditional least squares integration method with no compensation is demonstrated in Fig. 3(a), which shows large integration errors as the true surface does not satisfy the biquadratic spline assumption. The objective of the work is to reduce this type of error.

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 02/05/2014 Terms of Use: http://spiedl.org/terms

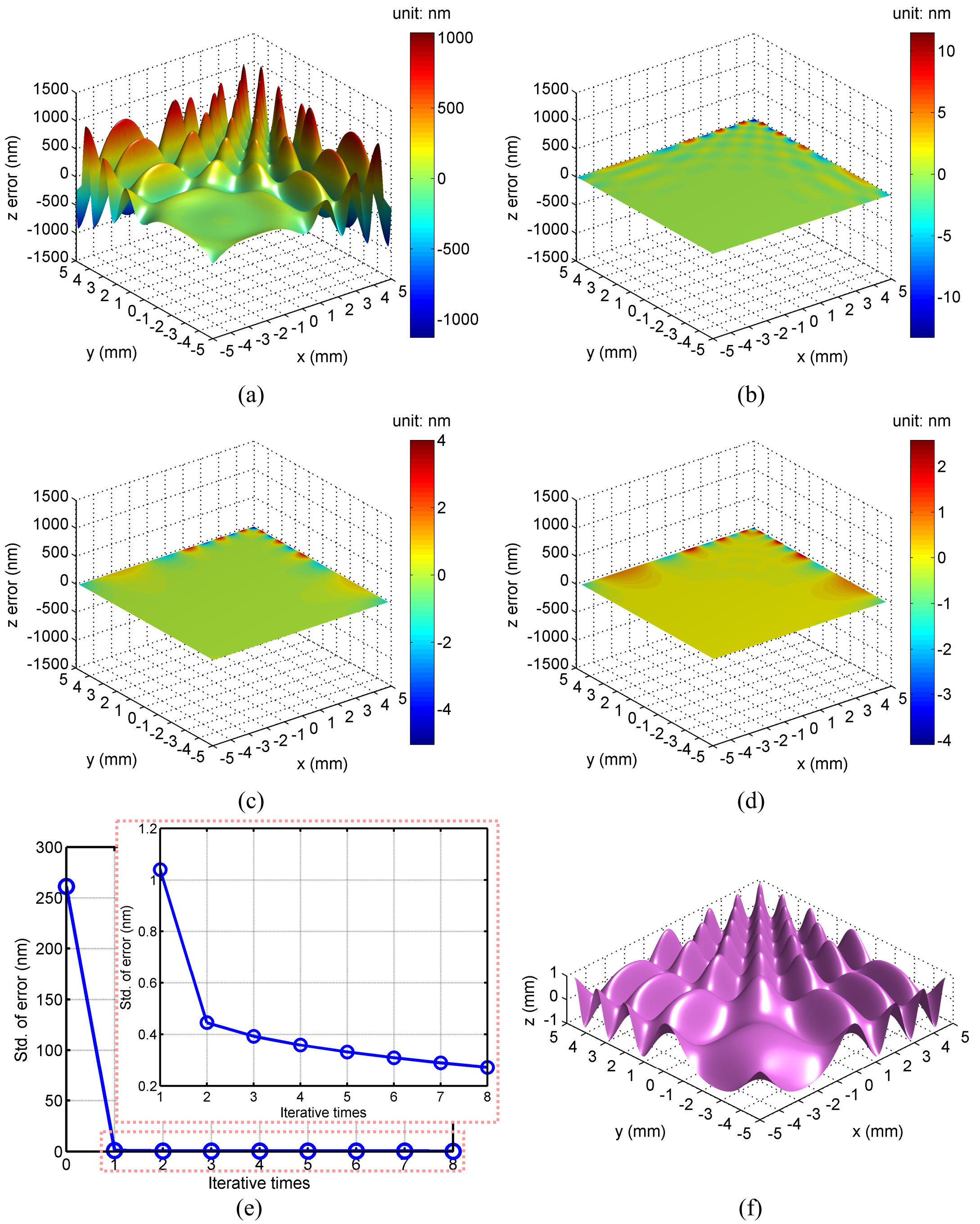


Fig. 3. Integration results. (a) Height error with no compensation, (b) after the 1st compensation, (c) after the 4th compensation, (d) after the 8th compensation, (e) standard deviations of height error after compensations, and (f) reconstructed result after the 8th compensation.

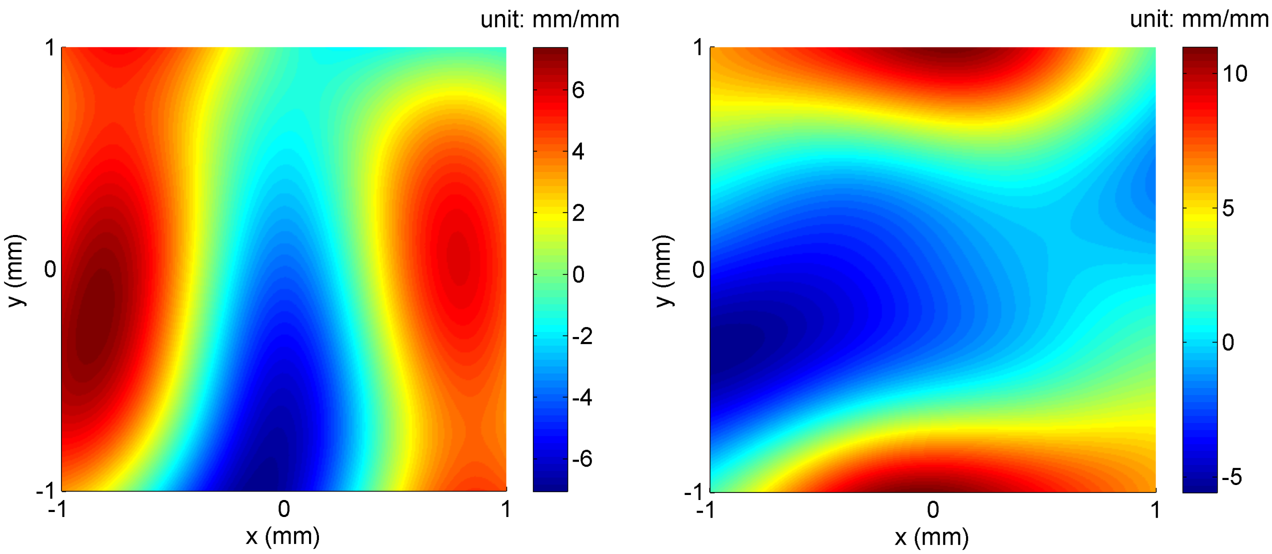
As shown in Fig. 3(b-d), with the proposed compensation process, height errors significantly reduced down to less than 1% as compared to Fig. 3(a). The majority of compensation is completed after the first one or two compensations, which is also shown in Fig. 3(d). After repeating the compensation with eight times, the out-of-plane height is accurately reconstructed as shown in Fig. 3(e).

The second example demonstrates the ability of the proposed method to operate in the presence of noise. The true height is defined by

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 02/05/2014 Terms of Use: http://spiedl.org/terms

$$z = 3(1-x)^2 \cdot e^{-x^2-(y+1)^2} - 10\left(\frac{x}{5} - x^3 - y^5\right) \cdot e^{-x^2-y^2} - \frac{1}{3} e^{-(x+1)^2-y^2}, \tag{10}$$

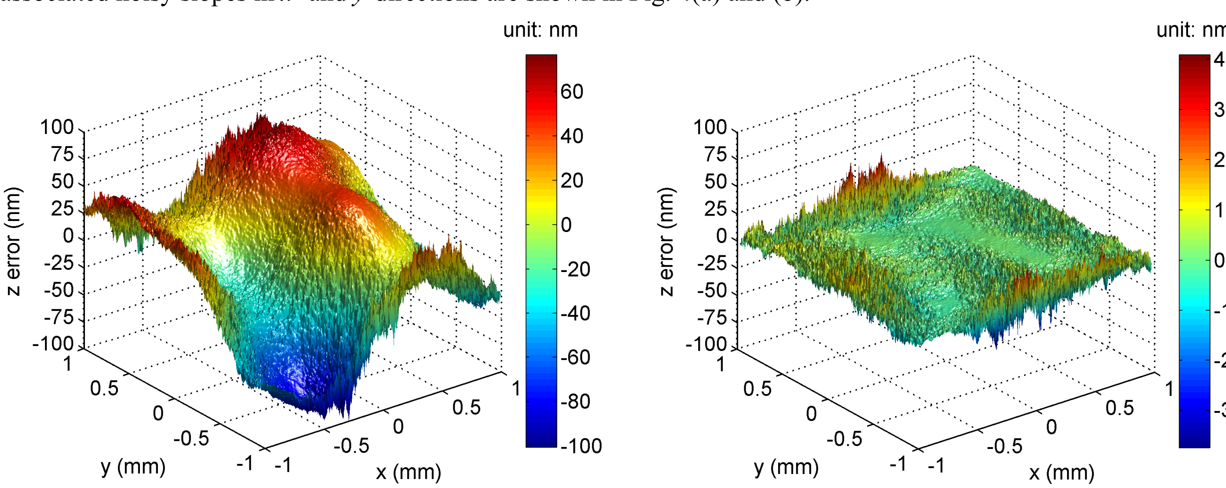
where the in-plane dimensions x and y are limited within a range from -1 to 1 with sampling points of 400 by 400, and the corresponding out-of-plane height varies by more than 5 mm.



(a) (b)

Fig. 4. Integrate gradient data with noise. (a) slope $p = \mathrm{d}z/\mathrm{d}x$, (b) slope $q = \mathrm{d}z/\mathrm{d}y$.

In practical measurement, noise is normally distributed on fringe phase, and furthermore can be approximately considered as normally distributed on surface normal directions. Normally distributed angular noise with a standard deviation of 8 arcsec is added in the normal directions, which is a typical noise level in fringe reflection technique. The associated noisy slopes in x- and y-directions are shown in Fig. 4(a) and (b).



(a) (b)

Fig. 5. Integration results. (a) Height error before compensation, (b) height error after iterative compensation, and (c).

The integration result from the traditional least squares method is shown in Fig. 5(a). Without any compensation, the height error is not only because of the influence from gradient noise but also the mismatch between the assumed biquadratic spline surface and the actual surface to be reconstructed. However, most of the errors due to this assumption can be compensated through the iterative process with the height error shown in Fig. 5(b).

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 02/05/2014 Terms of Use: http://spiedl.org/terms

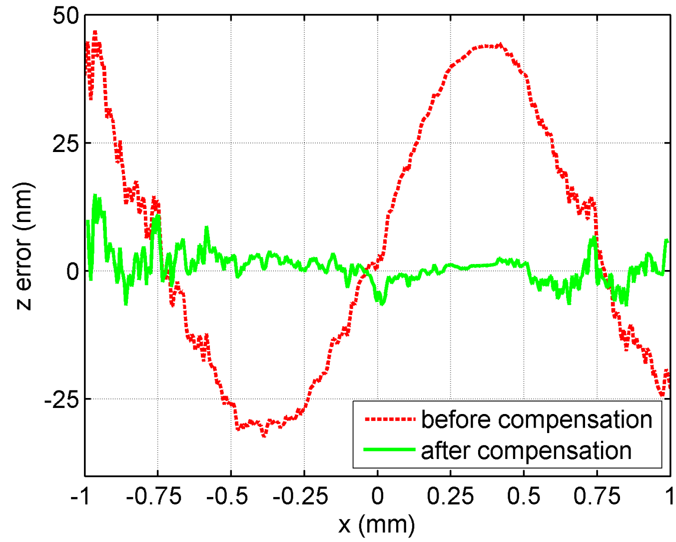


Fig. 6. Height error in one line (y=0)

A line profile (y=0) of errors before and after compensation is compared in Fig. 6, from which it can be easily found that the iterative process improves the integration accuracy with successful elimination of the waviness, and the residual error is mainly from the gradient noise. The result shows the iterative least squares integration method is able to reconstruct shapes from gradient data in the presence of noise.

4. Discussion

As noted in literature, Ettl's RBFs based method [10] is an accurate integration approach as well. A comparison of these two methods is also conducted to show the relative performances. As mentioned above, with no stitching process, the RBFs based integration method can only handle small piece of dataset because of its large memory consumption. Therefore the true height distribution is set as a small piece of spherical surface shown in Fig. 7(a) with

$$z = 90 - \sqrt{90^2 - x^2 - y^2}\,, \tag{11}$$

where the in-plane dimensions x and y vary from -4 to 4 with 40 by 40 sampling points such that the dataset can be handled by both methods. Normally distributed angular noise with a standard deviation of 8 arcsec is added onto the surface normal directions.

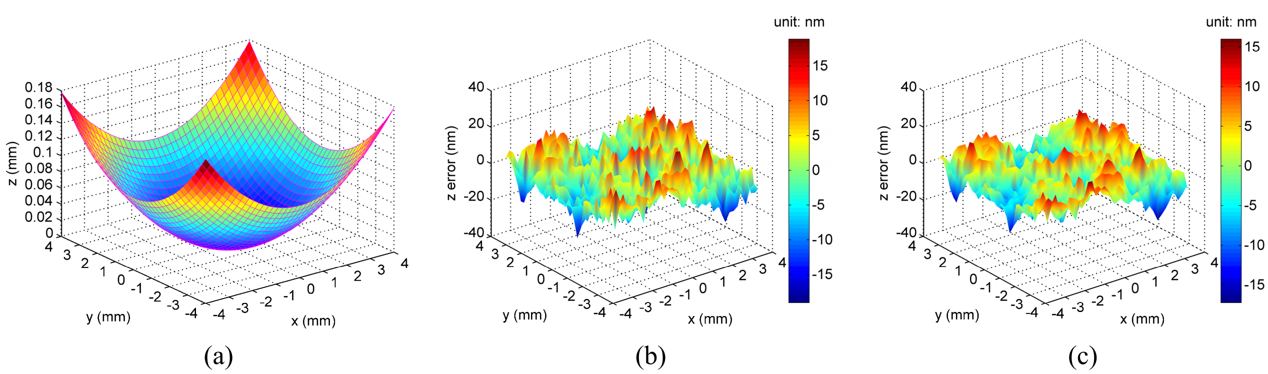


Fig. 7. Integration comparison with small dataset. (a) True shape, and (b-c) height error with the RBFs based integration method (b) and the proposed integration method (c).

Integration results of RBFs based method and the proposed method are shown with their height errors in Fig. 7(b) and (c), respectively. It is found that both methods are able to successfully reconstruct the shape from gradient data with a

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 02/05/2014 Terms of Use: http://spiedl.org/terms

comparable accuracy. Although least squares estimation is involved in both methods, there are some differences between them, especially when handling large datasets. The RBFs based integration method accurately integrates every subset in a sequence and then stitch all subsets together to get the entire shape. That means the heavy work is spatially distributed and completed in sequence. On the other hand, the proposed method works on the whole region or at least few relatively large blocks if blocking is necessary for some huge datasets. The issue of large memory occupation is solved by implementing simple integration models with sparse matrices and reusing the same memory during iterative calculation in time axis.

The comparison on accuracy and speed of these two methods is shown in Fig. 8. First, the accuracy is compared by integrating the same noisy gradient dataset with different sampling rates. The plot in Fig. 8(a) shows the standard deviations of height errors with different methods. It is obvious that the proposed method is more accurate. Also the larger the size of subset the more insensitive to noise as well as the stitching process the integration result is.

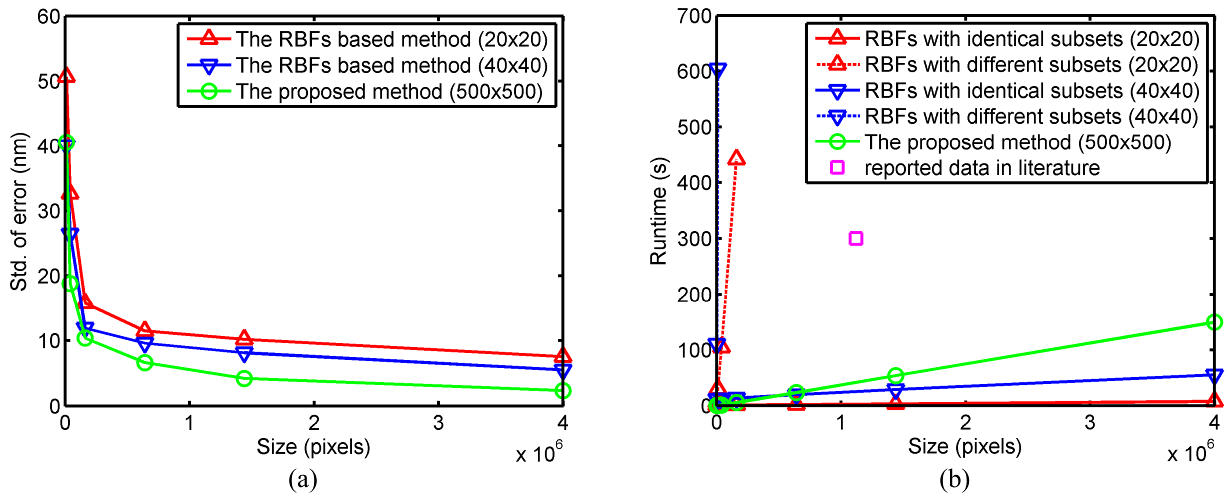


Fig. 8. Comparison of the RBFs method with the proposed method: accuracy (a) and speed (b).

Considering the necessary overlaps for stitching, the subset number is actually larger than the size ratio between the whole surface and the subset. The total runtime for integration is thus composed of the integration time in sequence and stitching time. It is also important to notice that if the subsets are identical for each patch, the integration time of the RBFs based method can be reduced by calculating the common matrix beforehand. Thus the RBFs based method is very fast when identical subsets are used as shown in Fig. 8(b) and also smaller subsets yields shorter runtime. However, in practical situation, some sampling data might be missing due to boundaries, unavailable regions, or poor quality, thus subsets are not identical and need to be recalculated, which increases the runtime as shown in Fig. 8(b). The proposed method copes with the complete and incomplete dataset in the same way and almost the same speed. Generally speaking the proposed method is fast, especially when the dataset can be directly handled without stitching. the proposed method runs faster, comparing to the reported runtime in Ref. [10] as around 300 seconds when reconstructing 10^6 surface values, which is also marked in Fig. 8(b). The limitation for this proposed method is investigated. As the traditional least squares integration method, the method can only handle complete or incomplete gradient dataset with rectangular mesh grid and it is not able to integrate arbitrarily distributed dataset, but the data are generally in rectangular mesh grid in many practical optical measurements.

5. Conclusion

This work attempts to improve the least squares integration method with iterative compensation to solve the issue of the incorrect biquadratic spline assumption. The proposed method is investigated by several simulations. Significant improvement in accuracy is verified by comparing with the traditional least squares integration method. Moreover, a comparison with the RBFs based integration method is carried out, and the results show the accuracy of both methods are comparable. The merits of the proposed method are accurate, fast, and able to handle large datasets. In summary, this least squares integration with iterative compensation method is an effective and accurate 2D integration tool to handle shape from slope problems in some gradient measuring based optical inspection applications.

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 02/05/2014 Terms of Use: http://spiedl.org/terms

Acknowledgements

The authors would like to thank Prof. Yuankun Liu in Sichuan University, China and Prof. Qian Kemao in School of Computer Engineering, Nanyang Technological University, Singapore, for many helpful discussions.

References

[1] Chen, F., Brown, G. M., and Song, M., "Overview of three-dimensional shape measurement using optical methods," Optical Engineering, 39(1), 10-22 (2000).

[2] Knauer, M. C., Kaminski, J., and Häusler, G., "Phase measuring deflectometry: a new approach to measure specular free-form surfaces," Proc. SPIE. 5457, 366-376 (2004).

[3] Bothe, T., Li, W., von Kopylow, C., and Jüptner, W. P. O., "High-resolution 3D shape measurement on specular surfaces by fringe reflection," Proc. SPIE. 5457, 411-422 (2004).

[4] Platt, B. C., and Shack, R., "History and principles of Shack-Hartmann wavefront sensing," Journal of Refractive Surgery, 17(5), S573-S577 (2001).

[5] Fried, D. L., "Least-square fitting a wave-front distortion estimate to an array of phase-difference measurements," J. Opt. Soc. Am., 67(3), 370-375 (1977).

[6] Hudgin, R. H., "Wave-front reconstruction for compensated imaging," J. Opt. Soc. Am., 67(3), 375-378 (1977).

[7] Hudgin, R. H., "Optimal wave-front estimation," J. Opt. Soc. Am., 67(3), 378-382 (1977).

[8] Southwell, W. H., "Wave-front estimation from wave-front slope measurements," J. Opt. Soc. Am., 70(8), 998-1006 (1980).

[9] Li, W., Bothe, T., von Kopylow, C., and Jüptner, W. P. O., "Evaluation methods for gradient measurement techniques," Proc. SPIE. 5457, 300-311 (2004).

[10] Ettl, S., Kaminski, J., Knauer, M. C., and Häusler, G., "Shape reconstruction from gradient data," Appl. Opt., 47(12), 2091-2097 (2008).

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 02/05/2014 Terms of Use: http://spiedl.org/terms