CrossMark

Spline based least squares integration for two-dimensional shape or wavefront reconstruction

Lei Huang[a,*], Junpeng Xue[a,b], Bo Gao[a,c,d], Chao Zuo[e], Mourad Idir[a]

[a] Brookhaven National Laboratory - NSLS II, 50 Rutherford Dr. Upton, NY 11973-5000, USA
[b] School of Aeronautics and Astronautics, Sichuan University, Chengdu 610065, China
[c] Shanghai Institute of Applied Physics, Chinese Academy of Sciences, Shanghai 201800, China
[d] University of Chinese Academy of Sciences, Beijing 100049, China
[e] Jiangsu Key Laboratory of Spectral Imaging & Intelligence Sense, Nanjing University of Science and Technology, Nanjing 210094, China

A R T I C L E   I N F O

A B S T R A C T

In this work, we present a novel method to handle two-dimensional shape or wavefront reconstruction from its slopes. The proposed integration method employs splines to fit the measured slope data with piecewise polynomials and uses the analytical polynomial functions to represent the height changes in a lateral spacing with the pre-determined spline coefficients. The linear least squares method is applied to estimate the height or wavefront as a final result. Numerical simulations verify that the proposed method has less algorithm errors than two other existing methods used for comparison. Especially at the boundaries, the proposed method has better performance. The noise influence is studied by adding white Gaussian noise to the slope data. Experimental data from phase measuring deflectometry are tested to demonstrate the feasibility of the new method in a practical measurement.

1. Introduction

Two-dimensional integration methods [1–4] are widely applied to reconstruct the height or wavefront from the measured gradient data in slope metrology, such as deflectometry [5,6] and wavefront sensing [7,8] etc.

The pioneer studies in two-dimensional integration can be found in wavefront reconstruction since the late 1970s [9–13]. Among these classical studies, Southwell's method [13] received great attention and success because of its good performance and simple implementation with the well-known Southwell geometry. It becomes the representative of the zonal integration methods. However, the integration accuracy is limited since it assumes the height distribution between two sampling points is only quadratic, which is not always true in reality. Based on this observation, an iterative compensation method was proposed to improve the accuracy [14]. By analyzing the Taylor theorem and truncation error, Li et al. proposed a straightforward method with higher order finite difference format [15], which is elegant and outperforms in a comparison [3] as it is more accurate than the traditional Southwell's method, and faster than the iterative method. Recently, Ren et al. proposed an easy implementation of Li's method for incomplete dataset or even in arbitrary domain [16].

In this work, we present a novel spline based least squares method

for two-dimensional shape or wavefront reconstruction from slopes in rectangular grids. Benefitted from high accuracy of spline fitting, the reconstruction accuracy can be improved. A comparative study with Southwell's method and Li's method is conducted in this work. The three methods share the same grid geometry (the Southwell geometry), as shown in Fig. 1. One of the beauties of this grid geometry is that the height reconstruction happens exactly at the same locations of slope measurement.

2. Principle

In the proposed method, the zonal relations of the neighboring height values are described as

$$
\begin{cases}
z_{m,n+1} - z_{m,n} = \sum_{k=0}^{3} \frac{1}{k+1} c^x_{m,n,k} \Delta x^{k+1}_{m,n} \\
z_{m+1,n} - z_{m,n} = \sum_{k=0}^{3} \frac{1}{k+1} c^y_{m,n,k} \Delta y^{k+1}_{m,n}
\end{cases}
\tag{1}
$$

where $\Delta x_{m,n} = x_{m,n+1} - x_{m,n}$ and $\Delta y_{m,n} = y_{m+1,n} - y_{m,n}$ are the x- and y-step sizes at matrix location (m, n) as show in Fig. 1. $c^x_{m,n,k}$ and $c^y_{m,n,k}$ are the coefficients of the kth order piecewise polynomials starting at (m, n), which are determined through the cubic spline fitting of the mth row of x-slopes and the nth column of y-slopes, respectively.

For instance, the x-slope at $(m, n+1)$ and y-slope at $(m+1, n)$ can be
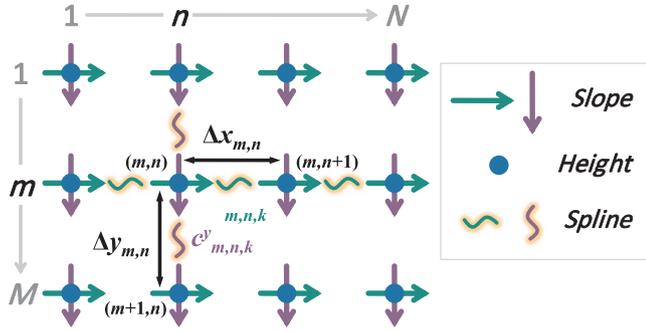
Fig. 1. In Southwell geometry, the height is reconstructed at the same locations where slopes are measured. In our method, slopes are fitted with cubic splines to represent height differences between spacing.

represented by piecewise polynomials starting at (m, n) as

$$s_{m,n+1}^x = \sum_{k=0}^{3} c_{m,n,k}^x \Delta x_{m,n}^k. \tag{2}$$

$$s_{m+1,n}^y = \sum_{k=0}^{3} c_{m,n,k}^y \Delta y_{m,n}^k. \tag{3}$$

The measured slopes and unknown height are consequently described with piecewise cubic and quartic polynomials, respectively. More significantly, slopes at boundaries of the dataset can be easily and accurately represented by setting the boundary condition of splines as the "natural boundary condition".

By integrating the analytical polynomial functions in Eq. (2) or Eq. (3) with the spline determined coefficients $c_{m,n,k}^x$ or $c_{m,n,k}^y$, the height difference after a lateral step $\Delta x_{m,n}$ or $\Delta y_{m,n}$ can be calculated through the right hand sides of Eq. (1). The linear least squares solution of height can be described as

$$\begin{bmatrix} z_{1,1} \\ z_{2,1} \\ \vdots \\ z_{M,N} \end{bmatrix} = (\mathbf{D}^\mathbf{T}\mathbf{D})^{-1}\mathbf{D}^\mathbf{T}\mathbf{G}, \tag{4}$$

where the symbol $(\cdot)^\mathbf{T}$ stands for the transpose operation, and $(\cdot)^{-1}$ is the matrix inverse. The sparse matrix $\mathbf{D}$ and vector $\mathbf{G}$ are

$$\mathbf{D} = \begin{bmatrix} -1 & 0 & \cdots & 0 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & -1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & 0 & -1 & 0 & \cdots & 0 & 1 \\ -1 & 1 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & -1 & 1 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & -1 & 1 \end{bmatrix}. \tag{5}$$

$$\mathbf{G} = \begin{bmatrix} \sum_{k=0}^{3} \frac{1}{k+1} c_{k,1,1}^x \Delta x_{1,1}^{k+1} \\ \sum_{k=0}^{3} \frac{1}{k+1} c_{k,2,1}^x \Delta x_{2,1}^{k+1} \\ \vdots \\ \sum_{k=0}^{3} \frac{1}{k+1} c_{k,M,N-1}^x \Delta x_{M,N-1}^{k+1} \\ \sum_{k=0}^{3} \frac{1}{k+1} c_{k,1,1}^y \Delta y_{1,1}^{k+1} \\ \sum_{k=0}^{3} \frac{1}{k+1} c_{k,2,1}^y \Delta y_{2,1}^{k+1} \\ \vdots \\ \sum_{k=0}^{3} \frac{1}{k+1} c_{k,M-1,N}^y \Delta y_{M-1,N}^{k+1} \end{bmatrix}. \tag{6}$$

3. Simulation

In order to illustrate the excellent performance of the proposed method, a two-dimensional cosine function with varying local frequencies $z = \cos(2\pi x^2/3000)\cdot \cos(2\pi y^2/3000)$ is selected as the Surface Under Test (SUT) to reconstruct as shown in Fig. 2(a). Its corresponding analytically derived x-slope and y-slope are shown in Figs. 2(b) and (c). We set x-unit the same as y-unit and named as "lateral unit," [l. u.], and z-unit is symbolled as [z. u.]. The in-plane coordinates are sampled as $x=1, 2\ldots, 256$ [l. u.] and $y=1, 2 \ldots, 256$ [l. u.]. The value of height ranges in ± 1 [z. u.], and both x-slope and y-slope range within ± 1 [z. u. / l. u.], so the Peak-To-Valley (PTV) of the slopes are 2 [z. u. / l. u.] for the simulated SUT.

Three integration methods (Southwell's method [13], Li's algorithm 1 in Ref. [15], and our spline-based method in this work) are applied to reconstruct height from the slopes in Fig. 2(b-c) for a comparison. All these methods share the same sparse matrix $\mathbf{D}$, which has the less memory cost and computing time for the matrix inverse operation comparing to other two algorithms in Ref. [15]. It is a big advantage in handling huge slope datasets. This is one of reasons why we compare these three methods.

The reconstruction errors are illustrated in Fig. 3. It indicates that these zonal methods make larger reconstruction errors in higher frequency regions. For comparison purposes, Southwell's method has the largest reconstruction error with its Root Mean Square (RMS) $=2.6\times10^{-2}$ [z. u.] and PTV $=0.19$ [z. u.]. Li's reconstruction method ends up with errors of RMS $=5.8\times10^{-3}$ [z. u.] and PTV $=0.17$ [z. u.] showing significant improvement compared with Southwell's method. Lastly, the proposed method outperforms the others with reconstruction errors of RMS $=9.6\times10^{-4}$ [z. u.] and PTV $=0.03$ [z. u.] only.

It is obvious that the proposed spline-based method has better estimation at regions with high-frequency variations. More significantly, splines have naturally good performance at dataset boundaries. In contrast, four neighboring slopes in one direction are always required in Li's algorithm 1 in Ref. [15] which cannot be satisfied at
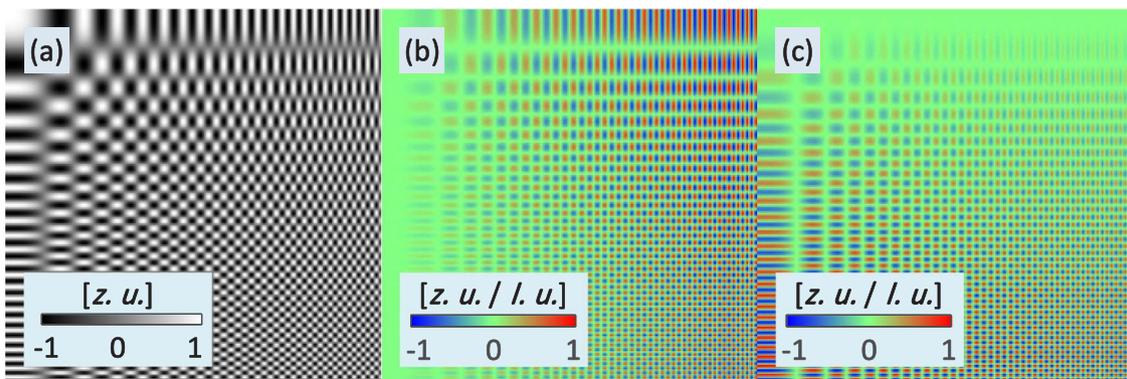
Fig. 2. A surface height (a) with varying local frequencies is chosen as the benchmark in simulation to test the performance of different methods in height reconstruction from x-slope (b), and y-slope (c).
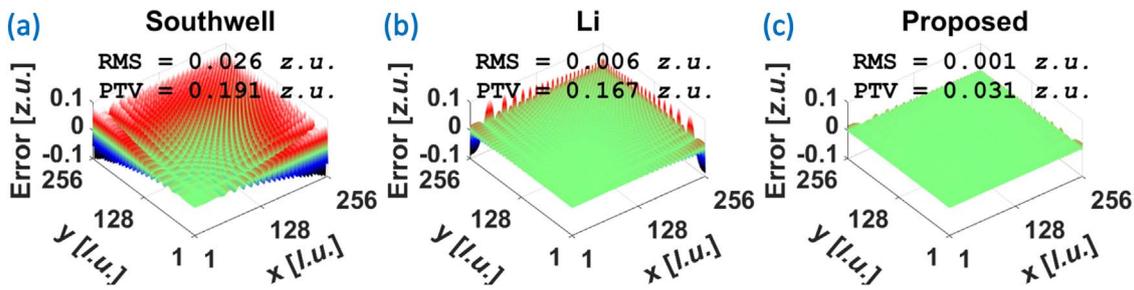
Fig. 3. The proposed method (c) has less errors comparing to the two existing methods: (a) Southwell's method, (b) Li's method. The codes in MATLAB® can be found in [17].

the boundaries.

Here we set the measuring range of the slope sensor as 5 times the slope PTV of the SUT in our simulation, which is a reasonable number in practical measurements. The simulated slopes PTV is 2 [z. u. / l. u.], so the slope measuring range is set as 10 [z. u. / l. u.] here. We define the Range-to-Precision Ratio (RPR) as the ratio of the slope measuring range of the sensor over its slope measuring precision. The RPR varies from 10^2 up to 10^5, which is a typical range for slope sensors in various applications. Additive white Gaussian noises with different standard deviations according to different RPR values are simulated into slopes to study the behavior of integrators under noise conditions.(Fig. 4).

These three methods offer very similar results when the RPR is only 10^2. When the RPR is approaching 10^3, the noise still dominates the total reconstruction error, but it can be noticed that the errors from Li's method and the proposed method are getting smaller than Southwell's result. When the RPR is higher than 10^4, the truncation error in numerical calculation especially in the high frequency region makes the major contribution to the total error for Southwell's and Li's methods. The proposed spline-based method performs better comparing to the other two methods in this range. It indicates an accurate integration method becomes more important when measuring high-frequency surface variations by using a higher precision slope metrology tool, such as the stitching shack-Hartmann optical head [18] which typically has a RPR value about 10^5.

Moreover, the proposed method can be used to handle incomplete slope datasets. In this condition, the spline fitting is implemented section by section instead of an entire vector, due to the segmentation by invalid regions with no slope available. At the boundaries of the invalid region, the cubic spline has its good natural solution once at least 4 valid pixels are connected in one direction. If there are only 3 or 2 slope values in one section, a lower order spline will be used in fitting those slopes or Southwell-derived **G** values can become a backup.

As illustrated in Fig. 5, some of the slope values are removed from the slopes in Fig. 2, which is common in real measurements due to the unacceptable surface quality. The performance of the proposed spline-based integration method is tested with this type of incomplete slope dataset. The RPR of the measurement system is set as 10^4 to have certain slope noise in the simulation. The proposed method can accurately reconstruct the surface as shown in Fig. 6(a). Of course, the regions with no slope data cannot be reconstructed in principle. The reconstruction error in Fig. 6(b) is only 1×10^{-3} *z. u.* RMS and 1.4×10^{-2} *z. u.* PTV, comparing to the reconstruction errors of Southwell's method (1.2×10^{-2} *z. u.* RMS and 9.8×10^{-2} *z. u.* PTV) in Fig. 6(c) and Li's method (5×10^{-3} *z. u.* RMS and 5.7×10^{-2} *z. u.* PTV) in Fig. 6(d).
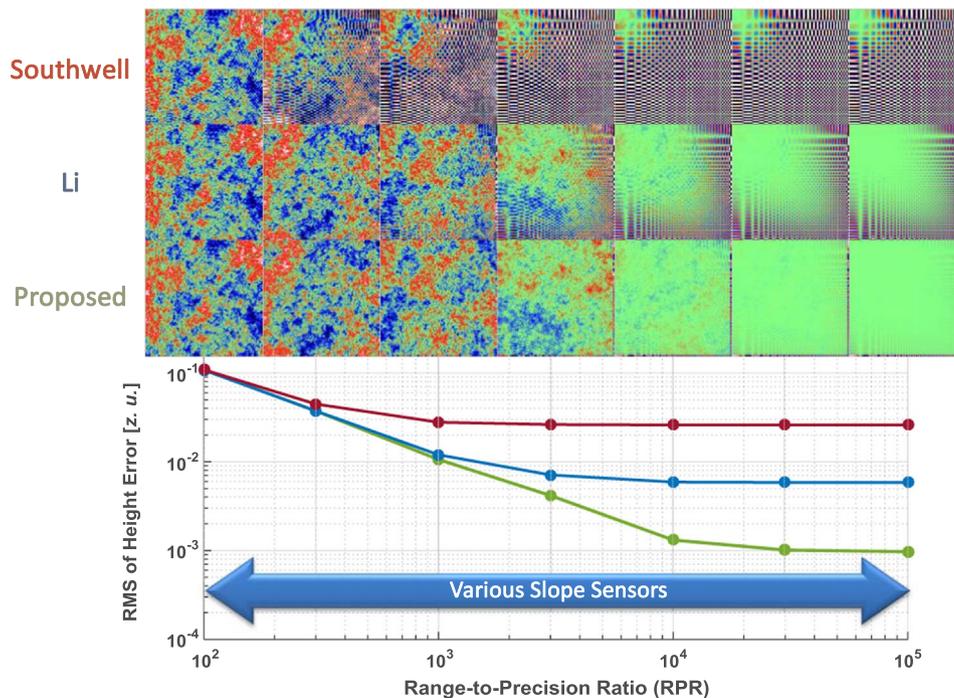
Fig. 4. While the RPR of the slope sensor increases (slope noise decreasing with a constant measuring range in our simulation) from 10^2 to 10^5, the error owing to the slope noise dominates the reconstruction error at first and then gradually has less impact comparing to the truncation error of algorithms. When RPR is higher than 10^4, the reconstruction accuracy of the proposed method is much better than the other methods in comparison, especially in high frequency regions.
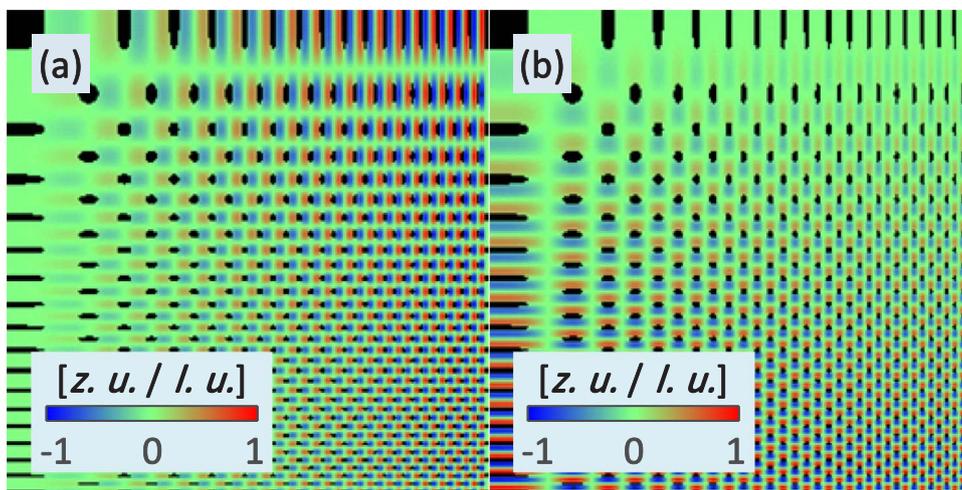
Fig. 5. Incomplete *x*-slope (a) and *y*-slope (b) data are simulated to study the performance of the proposed method.

4. Experiment

To demonstrate the feasibility of the proposed method in an actual measurement, experimental slope results from phase measuring deflectometry are used as a data sample for reconstruction. The SUT is about 15 mm × 15 mm with slopes ranging in ± 17 mrad. The slopes are integrated by the proposed method as illustrated in Fig. 7.

Although there are many invalid data in the measured slopes and the aperture shape is not rectangular, the proposed method still can successfully reconstruct the surface height as shown in Fig. 7(c). The difference of two reconstructions (the proposed method vs. Li's method) is illustrated in Fig. 7(d). The differences around the edges of the bumps on the surface are relatively bigger due to the rapid changes in slopes

locally.

5. Discussion

Since the spline fitting of slopes takes additional time, the speed of the proposed spline-based integration method is slower than the other two zonal methods, but their speed difference is insignificant. Fig. 8 compares the time of using these three methods for different slope data size. The comparison is carried out in MATLAB® running on an i7-4600M CPU @ 2.90 GHz with 16 GB memory. Generally speaking, the proposed method takes 30–80% longer time than Li's method does depending on the data size. The larger size yields the lower ratio of their speed difference.
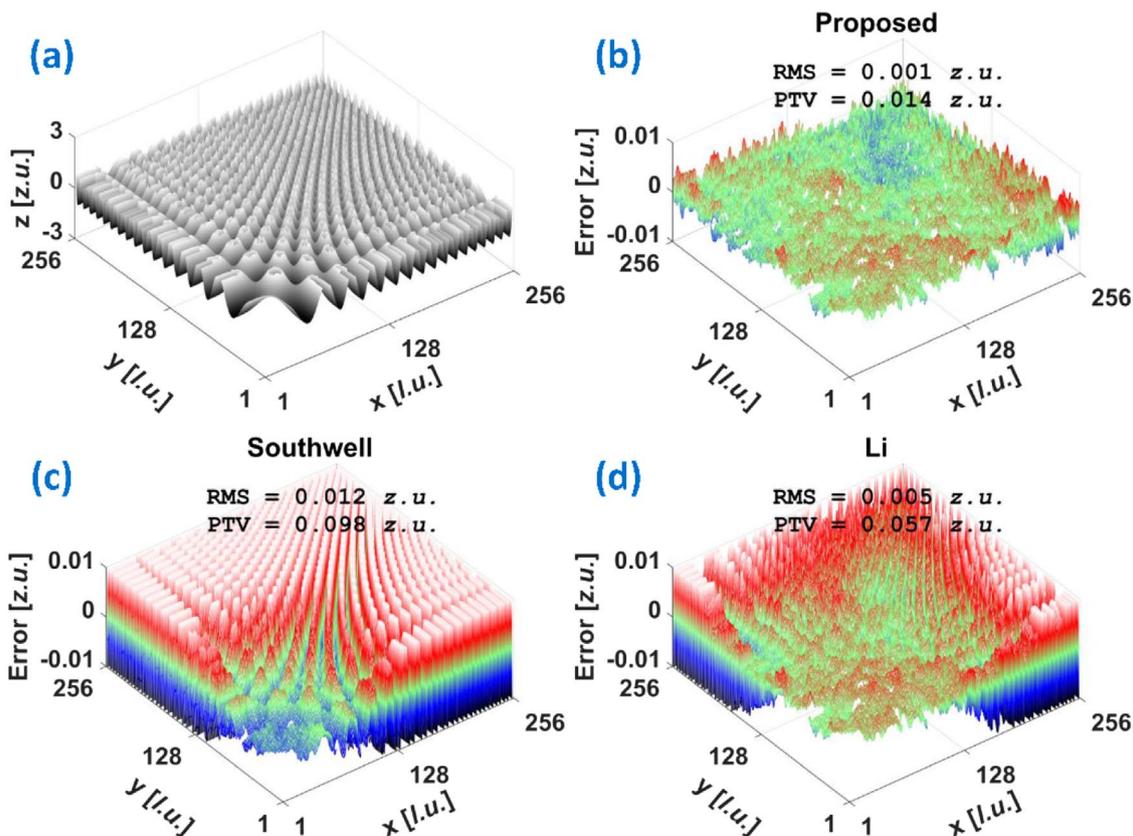
Fig. 6. The SUT can be reconstructed by the proposed method even if there are invalid slope regions (a). The error of the proposed reconstruction (b) is smaller than those of the Southwell's method (c) and Li's method (d). The same axes limits and color map are set in (b)-(d) to highlight their difference.
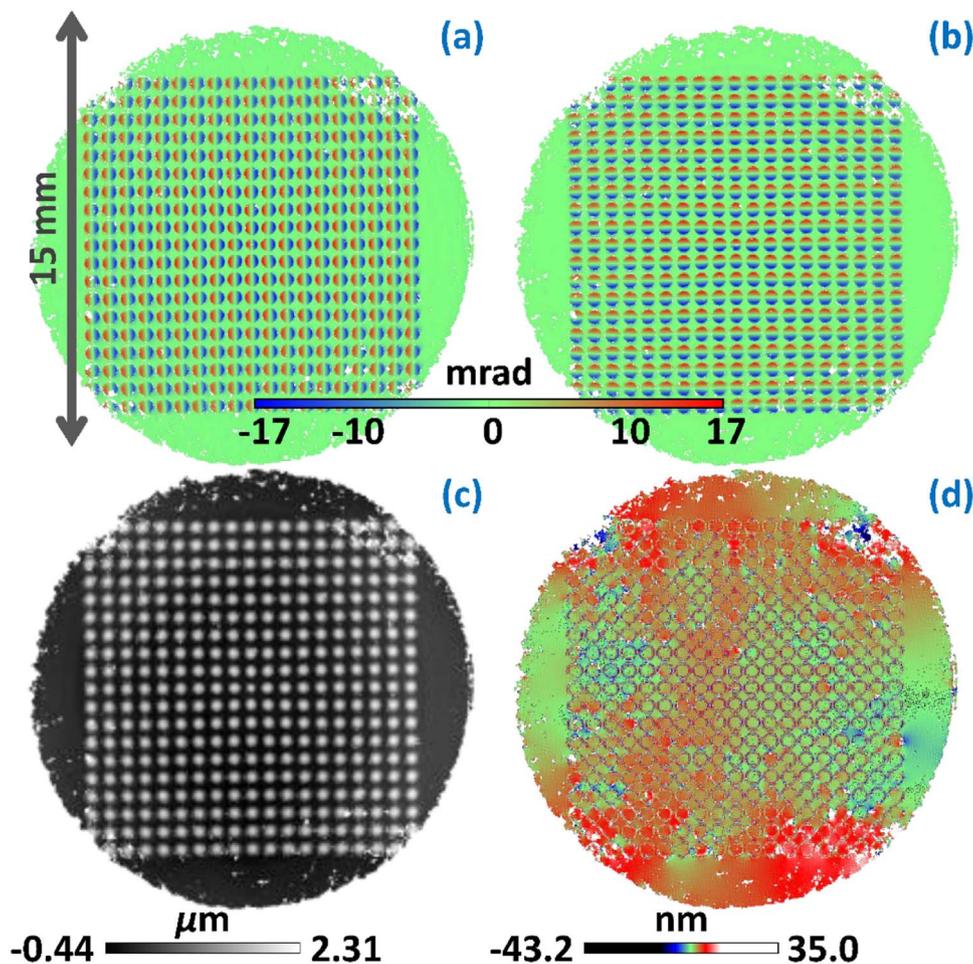
Fig. 7. Experimental *x*-slope (a) and *y*-slope (b) from deflectometry are integrated by using the proposed spline-based method, and the resultant height map (c) has a reconstruction difference (d) comparing to Li's method.
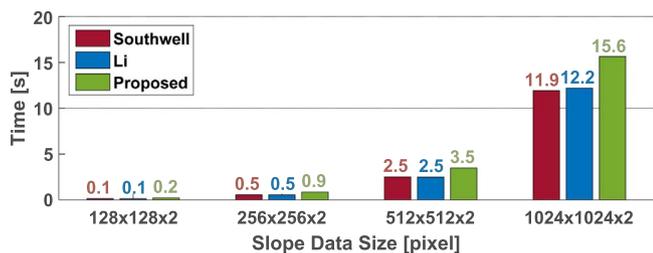
Fig. 8. With the slope data size increasing, the time for integration gets longer for all the methods. The proposed method takes slightly longer time comparing to the other two methods.

The new method is proposed to deal with slope data in rectangular grids, and it has an intrinsic limitation to process the slopes in quadrilateral grids, or the even more general triangular grids. This can be a potential work for future.

6. Conclusion

In summary, we present a novel two-dimensional integration method in this work to reconstruct the height or wavefront from its sampled first derivatives (the measured slopes). The well-established spline technique is employed to fit the slopes with piecewise polynomials. The analytical polynomial functions with determined coefficients are employed to calculate the height variations for each spatial step in *x*-and *y*-directions. The final height result is estimated with the linear least squares method. The codes in MATLAB® can be downloaded

from [17]. A comparison with two similar existing integration techniques and the noise influence are investigated through simulation to reveal the advantages of the proposed method. The proposed method show great performance at the boundaries of the dataset or holes. Working with the experimental deflectometry slope data demonstrates the proposed method is feasible and effective in a practical integration.

References

[1] Ares M, Royo S. Comparison of cubic B-spline and Zernike-fitting techniques in complex wavefront reconstruction. Appl Opt 2006;45:6954–64.
[2] Huang L, Asundi AK. Framework for gradient integration by combining radial basis functions method and least-squares method. Appl Opt 2013;52:6016–21.
[3] Huang L, Idir M, Zuo C, Kaznatcheev K, Zhou L, Asundi A. Comparison of two-dimensional integration methods for shape reconstruction from gradient data. Opt Lasers Eng 2015;64:1–11.
[4] Dai F, Tang F, Wang X, Sasaki O, Feng P. Modal wavefront reconstruction based on Zernike polynomials for lateral shearing interferometry: comparisons of existing algorithms. Appl Opt 2012;51:5028–37.
[5] Ettl S, Kaminski J, Knauer MC, Häusler G. Shape reconstruction from gradient data. Appl Opt 2008;47:2091–7.
[6] Ren H, Gao F, Jiang X. Least-squares method for data reconstruction from gradient

data in deflectometry. Appl Opt 2016;55:6052–9.

[7] Mochi I, Goldberg KA. Modal wavefront reconstruction from its gradient. Appl Opt 2015;54:3780–5.

[8] Seifert L, Tiziani HJ, Osten W. Wavefront reconstruction with the adaptive Shack–Hartmann sensor. Opt Commun 2005;245:255–69.

[9] Fried DL. Least-square fitting a wave-front distortion estimate to an array of phase-difference measurements. J Opt Soc Am 1977;67:370–5.

[10] Hudgin RH. Wave-front reconstruction for compensated imaging. J Opt Soc Am 1977;67:375–8.

[11] Hudgin RH. Optimal wave-front estimation. J Opt Soc Am 1977;67:378–82.

[12] Noll RJ. Phase estimates from slope-type wave-front sensors. J Opt Soc Am 1978;68:139–40.

[13] Southwell WH. Wave-front estimation from wave-front slope measurements. J Opt Soc Am 1980;70:998–1006.

[14] Huang L, Asundi A. Improvement of least-squares integration method with iterative compensations in fringe reflectometry. Appl Opt 2012;51:7459–65.

[15] Li G, Li Y, Liu K, Ma X, Wang H. Improving wavefront reconstruction accuracy by using integration equations with higher-order truncation errors in the Southwell geometry. J Opt Soc Am A 2013;30:1448–59.

[16] Ren H, Gao F, Jiang X. Improvement of high-order least-squares integration method for stereo deflectometry. Appl Opt 2015;54:10249–55.

[17] Huang L, MATLAB codes for spline based least squares integration for two dimensional shape or wavefront reconstruction, ⟨https://github.com/huanglei0114/Spline-based-least-squares-integration-for-two-dimensional-shape-or-wavefront-reconstruction⟩.

[18] Idir M, Kaznatcheev K, Dovillaire G, Legrand J, Rungsawang R. A 2 D high accuracy slope measuring system based on a stitching Shack Hartmann optical head. Opt Express 2014;22:2770–81.