Mathematical construction and perturbation analysis of Zernike discrete orthogonal points

Zhenguang Shi,* Yongxin Sui, Zhenyu Liu, Ji Peng, and Huaijiang Yang

State Key Laboratory of Applied Optics, Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, Jilin 130033, China

*Corresponding author: shizg@ciomp.ac.cn

Received 20 March 2012; revised 28 April 2012; accepted 29 April 2012; posted 30 April 2012 (Doc. ID 165073); published 18 June 2012

Zernike functions are orthogonal within the unit circle, but they are not over the discrete points such as CCD arrays or finite element grids. This will result in reconstruction errors for loss of orthogonality. By using roots of Legendre polynomials, a set of points within the unit circle can be constructed so that Zernike functions over the set are discretely orthogonal. Besides that, the location tolerances of the points are studied by perturbation analysis, and the requirements of the positioning precision are not very strict. Computer simulations show that this approach provides a very accurate wavefront reconstruction with the proposed sampling set. © 2012 Optical Society of America

OCIS codes: 080.1005, 010.7350, 220.4840.

1. Introduction

In optical testing, Zernike functions have been widely used to describe optical surfaces or wavefronts, since they are made up of polynomial terms that are of the same form as a Seidel aberration [1,2]. The surface aberrations or wavefronts of an optical system are recorded with a uniform sampling array like a CCD and then recovered with certain algorithms, such as phase shifting algorithms [3,4] in modern digital phase shifting interferometers. In optomechanical analysis, predicting optical performance under the actual operational environment of an optical system often requires importing finite element computed surface displacements into the optical model such as the Zernike description of surface undulation [5,6]. The finite element meshing for optics is carried out with certain types of grid partitioning.

In the preceding cases, Zernike polynomials' expansion coefficients are all required to calculate from the discrete data over some types of sampling points. As is known to us, Zernike polynomials are orthogonal within the unit circle, but they are not the case over the commonly used discrete points unless they are uniformly distributed and dense enough [7]. Fortunately, by studying the mathematical properties of Zernike functions, Pap and Schipp have designed a sampling fashion of the unit circle elaborately, which ensures that the Zernike polynomials are discrete orthogonal over the constructed set mathematically [8], which can be used to deal with the preceding engineering problem. But they do not analyze the locating tolerance of sampling points since the actual sampling points will not coincide with the ideal ones exactly in practice, such as CCD equidistant sampling or mesh points in finite element analysis. The locating errors will be studied by perturbation analysis in our work.

The paper is organized as follows: First the mathematical construction method of the discrete sampling points within the unit circle is introduced in detail in Section 2. And then in Section 3 the perturbation analysis about the location errors of the actual sampling points is studied. Computer simulations in Section 4 are to validate the proposed sampling method experimentally. The conclusion about the possible application of the proposed method is summarized in Section 5.

1559-128X/12/184210-05$15.00/0

2. Mathematical Construction

Zernike functions are variable separable in the polar coordinates

$$Z_n^m(\rho,\theta) = \sqrt{n+1}R_n^m(\rho)e^{\mathrm{i}m\theta}, \tag{1}$$

the orthogonality relation for radial polynomials is

$$\int_0^1 R_n^m(\rho)R_{n'}^{m'}(\rho)\rho\mathrm{d}p = \frac{\delta_{nn'}}{2(n+1)}, \tag{2}$$

and the azimuthal polynomials are orthogonal complex triangle functions satisfying

$$\int_0^{2\pi} e^{\mathrm{i}m\theta}\overline{e^{\mathrm{i}m'\theta}}\mathrm{d}\theta = 2\pi\delta_{mm'}. \tag{3}$$

That is to say that Zernike functions are orthogonal within the unit circle

$$\frac{1}{\pi}\int_0^{2\pi}\int_0^1 Z_n^m(\rho,\theta)\overline{Z_{n'}^{m'}(\rho,\theta)}\rho\mathrm{d}\rho\mathrm{d}\theta = \delta_{nn'}\delta_{mm'}. \tag{4}$$

The mathematical construction method of the discrete sampling points within the unit circle is introduced as follows [8]: Let us denote the roots of Legendre polynomials $P_N(x)$ of order N by $\lambda_k \in (-1,1)$, $k \in 1,\ldots,N$. Constructing the Lagrange interpolation fundamental polynomials

$$l_k(x) = \prod_{\substack{j=1\\ j\neq k}}^{N}\frac{x-\lambda_j}{\lambda_k-\lambda_j}, \tag{5}$$

and calculating the corresponding integral numbers

$$A_k = \int_{-1}^{1} l_k(x)\mathrm{d}x, \tag{6}$$

then for every polynomial $f(x)$ of order less than $2N$, the following relation is satisfied:

$$\int_{-1}^{1} f(x)\mathrm{d}x = \sum_{k=1}^{N} f(\lambda_k)A_k. \tag{7}$$

So for Zernike radial polynomials, define

$$\rho_k = \sqrt{\frac{1+\lambda_k}{2}} \in (0,1), \quad k = 1,\ldots,N, \tag{8}$$

if

$$\begin{aligned}\frac{n+|m|}{2}+\frac{n'-|m'|}{2} &\le 2N-1,\\ \frac{n-|m|}{2}+\frac{n'+|m'|}{2} &\le 2N-1,\end{aligned} \tag{9}$$

then

$$\begin{aligned}\frac{\delta_{nn'}}{2(n+1)} &= \int_0^1 R_n^m(\rho)R_{n'}^{m'}(\rho)\rho\mathrm{d}\rho\\ &= \frac{1}{4}\sum_{k=1}^{N} R_n^m(\rho_k)R_{n'}^{m'}(\rho_k)A_k.\end{aligned} \tag{10}$$

And for Zernike azimuthal functions, because of their orthogonality, just define enough uniform sampling points over the circle

$$\theta_j = \frac{2\pi j}{4N+1}, \quad j = 0,\ldots,4N, \tag{11}$$

and then azimuthal functions are orthogonal over these points.

In conclusion, the set of nodal points can be constructed as follows:

$$X = \{z_{jk} = (\rho_k,\theta_j), k = 1,\ldots,N, j = 0,\ldots,4N\}. \tag{12}$$

Sort double-index Zernike polynomials by some ordering type $p = p(m,n)$ such as the 36-term Fringe Zernike polynomials commonly used in optical testing. Then if the orthogonality condition in Eq. (9) is founded, the coupling coefficient of the pth and p'th Zernike functions over the discrete sampling points is

$$\begin{aligned}G_{pp'} &= \sum_{k=1}^{N}\sum_{j=0}^{4N} Z_n^m(\rho_k,\theta_j)\overline{Z_{n'}^{m'}(\rho_k,\theta_j)}\frac{A_k}{2(4N+1)}\\ &= \frac{1}{\pi}\int_0^{2\pi}\int_0^1 Z_n^m(\rho,\theta)\overline{Z_{n'}^{m'}(\rho,\theta)}\rho\mathrm{d}\rho\mathrm{d}\theta\\ &= \delta_{nn'}\delta_{mm'}.\end{aligned} \tag{13}$$

So the Gram matrix $G = [G_{pp'}]$ formed by the coupling coefficients will be a unity matrix. See Fig. 1(b) for its deviations from the unity matrix I (called "orthogonality deviations" from now on). The orthogonality deviations of the 10^{-13} level imply that the Gram matrix is certain to be a unity matrix except for computer precision.

There is also an improvement about the number of points in the azimuthal direction. If Eq. (9) is founded, then $m_{\max} = N-1$ and $n_{\max} = 2(N-1)$. So to ensure the orthogonality of azimuthal functions, just $2m_{\max}+1 = 2N-1$ sampling points are sufficient. The azimuthal points are

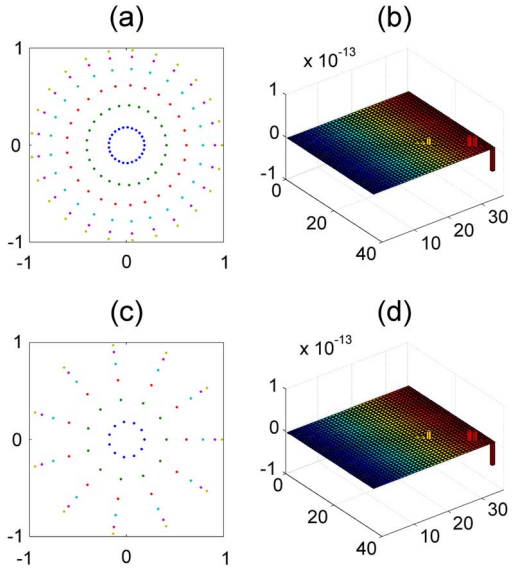

Fig. 1. (Color online) Sampling patterns and orthogonality deviations comparisons for $4N+1$ and $2N-1$ azimuthal points. (a) $4N+1$ azimuthal sampling, (b) orthogonality deviations of $4N+1$ sampling, (c) $2N-1$ azimuthal sampling, and (d) orthogonality deviations of $2N-1$ sampling.

$$\theta_j = \frac{2\pi j}{2N-1}, \quad j = 0, \ldots, 2N-2. \tag{14}$$

The comparison between the two sampling patterns and orthogonality deviations are illuminated in Fig. 1. One can see that the $2N-1$ sampling (66 points in total for $N=6$) is just enough to guarantee the orthogonality of Zernike polynomials without the use of the $4N+1$ sampling (150 points in total for $N=6$).

Then the Zernike reconstruction coefficients of wavefront $T(\rho,\theta)$ over the constructed discrete sampling points

$$C_{mn} = \sum_{k=1}^{N}\sum_{j=0}^{N'-1} T(\rho_k,\theta_j)\overline{Z_n^m(\rho_k,\theta_j)}\frac{A_k}{2N'} \times (N' = 2N-1 \,\&\, 4N+1) \tag{15}$$

will be absolutely accurate, because Zernike terms do not correlate with each other at all for their complete orthogonality.

3. Perturbation Analysis

As we have said before, the actual sampling points cannot coincide exactly with the mathematically constructed ideal ones in practice. In such cases, the orthogonality of Zernike polynomials in Eq. (13) is not founded exactly and will be perturbed slightly. This situation is analyzed as follows: without loss of generality, suppose that only a tiny perturbation with λ_1 changed to be $\lambda_1' = \lambda_1 + \Delta\lambda_1$. Then the Lagrange interpolation fundamental polynomials in Eq. (5) will be changed to

$$l_1'(x) = \prod_{j=2}^{N}\frac{x-\lambda_j}{\lambda_1'-\lambda_j} = \prod_{k=2}^{N} p_k \cdot l_1(x),$$

$$l_k'(x) = \frac{x-\lambda_1'}{\lambda_k-\lambda_1'}\cdot\prod_{\substack{j=2\\ j\neq k}}^{N}\frac{x-\lambda_j}{\lambda_k-\lambda_j} = p_k l_k(x) - \Delta\lambda_1 p_k \frac{l_k(x)}{x-\lambda_1}, \tag{16}$$

where

$$p_k = \frac{\lambda_1-\lambda_k}{\lambda_1'-\lambda_k}, \tag{17}$$

and the integral coefficients in Eq. (6) will be

$$A_1' = \int_{-1}^{1} l_1'(x)\mathrm{d}x = \prod_{k=2}^{N} p_k \cdot A_1,$$

$$A_k' = \int_{-1}^{1} l_k'(x)\mathrm{d}x = p_k A_k - \Delta\lambda_1 p_k B_k, \tag{18}$$

where

$$B_k = \int_{-1}^{1}\frac{l_k(x)}{x-\lambda_1}\mathrm{d}x, \tag{19}$$

so the discrete sum in Eq. (7) will be

$$\begin{aligned}\sum_{k=1}^{N} f(\lambda_k^1)A_k' &= f(\lambda_1')A_1' + \sum_{k=2}^{N} f(\lambda_k)A_K' \\ &\approx (f(\lambda_1) + \Delta\lambda_1 f'(\lambda_1))A_1' + \sum_{k=2}^{N} f(\lambda_k)A_K' \\ &= (f(\lambda_1)A_1 + \Delta\lambda_1 f(\lambda_1')A_1)\cdot\prod_{k=2}^{N} p_k \\ &\quad + \sum_{k=2}^{N} f(\lambda_k)p_k A_k - \Delta\lambda_1\cdot\sum_{k=2}^{N} f(\lambda_k)p_k B_k.\end{aligned} \tag{20}$$

Then the perturbation quantity of the discrete sum is

$$\begin{aligned}\Delta G &= \sum_{k=1}^{N} f(\lambda_k')A_k' - \sum_{k=1}^{N} f(\lambda_k)A_k \\ &\approx \Delta\lambda_1\left(\sum_{k=2}^{N}\left(\frac{f(\lambda_1)A_1 + f(\lambda_k)A_k}{\lambda_1-\lambda_k} + f(\lambda_k)B_k\right) - f(\lambda_1')A_1\right).\end{aligned} \tag{21}$$

One can see that for Zernike radial functions, when there is small perturbation of actual sampling points about the mathematically constructed ideal ones, the coupling coefficients are also perturbed very slightly.

But as pointed out in Eq. (21), the perturbation quantity of Zernike coefficient satisfies

$$\Delta C_{mn} = O(\Delta\lambda_1)(\Delta\lambda_1 = \lambda_1' - \lambda_1 \to 0). \qquad (22)$$

That is to say that it is just an infinitesimal of the same order as $\Delta\lambda_1$ in a first-order approximation that can almost be negligible. The analogous conclusion can be achieved for Zernike azimuthal functions too, which is that the perturbation quantity is also an infinitesimal of the same order as $\Delta\theta_1$. So as long as the actual sampling points do not deviate from the ideal ones largely, the reconstruction coefficients of Zernike polynomials will not deviate from the real ones seriously.

4. Computer Simulations

To demonstrate the preceding viewpoint experimentally, one can perturb the constructed ideal points randomly within a tiny quantity via computer simulations. Then the Gram matrix $G = [G_{pp'}]$ in Eq. (13) is recalculated to compare with the ideal one, and the orthogonality deviations between G and I are plotted for observation.

For the forward 36-term Fringe Zernike polynomials, $m_{\max} = 5$ and $n_{\max} = 10$. So to ensure their orthogonality, only $N = 6$ radial points and $2N - 1 = 11$ azimuthal points (total of 66 points) are required. See Figs. 1(c) and 1(d) for the sampling patterns and corresponding orthogonality deviations.

The first result is about the radial coordinate perturbation. The roots of Legendre polynomials are taken four decimal places; that is to say that the errors of roots lie in the $[-5, 5) \times 10^{-4}$ region. The orthogonality deviations resulting from the tiny perturbation are illustrated in Fig. 2(b). Then the azimuthal coordinates are perturbed by a random variable subject to uniform distribution, which lies in $[-0.01, 0.01]^\circ$, that is, the $[-1.75, 1.75] \times 10^{-4}$ rad. The corresponding orthogonality deviations are shown in Fig. 2(c). At last the perturbations in both directions are studied together, and the orthogonality deviations are shown in Fig. 2(d), where the radial and azimuthal perturbation quantities are as much as that used in Fig. 2(b) and in Fig. 2(c), respectively. It is seen that the differences of all the coupling coefficients are not more than the 10^{-3} level, which supports our argument in Eq. (21) that the perturbation quantities of the coupling coefficients are just mathematical infinitesimals of the same order as the perturbation of coordinates that are of the 10^{-4} level.

It is shown below that the supposed perturbation quantities of the radial and azimuthal directions are reasonable physically. For $N = 6$ radial points, the discrete radial positions are $\rho \approx [0.1837, 0.4116, 0.6170, 0.7870, 0.9114, 0.9830]$, respectively. So given a 4-in. mirror as an example, the strictest requirement of positioning precision is at the first radial point ρ_1, where the requirements are $50\ \text{mm} \times 0.1837 \times 5 \times 10^{-4} = 4.6\ \mu\text{m}$ and

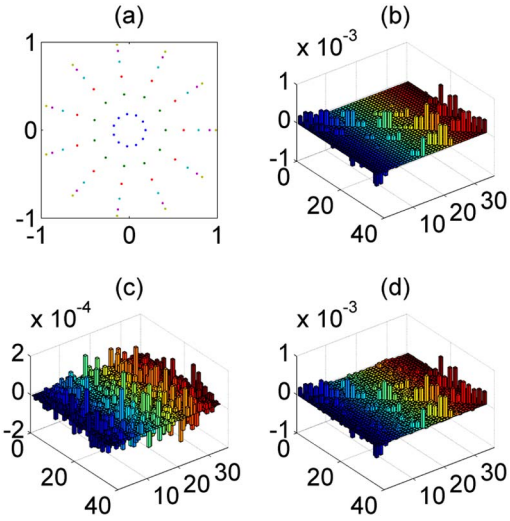

Fig. 2. (Color online) Sampling patterns and orthogonality deviation comparisons for radial and azimuthal perturbation about ideal points. (a) $2N - 1$ sampling patterns, (b) deviations of 10^{-4} levels in radial direction, (c) deviations of 10^{-4} levels in azimuthal direction, and (d) deviations of 10^{-4} levels in both directions.

$50\ \text{mm} \times 0.1837 \times 1.75 \times 10^{-4} = 1.6\ \mu\text{m}$, respectively, in the two directions. In the precise optomechanical analysis, the quantities of displacements of finite element mesh grids under a real operational environment will be significantly less than the preceding requirements. Also in the low frequency form metrology, the perturbation quantities are far less than the physical size per pixel. That is to say that the supposed perturbation quantities lie in the appropriate regions certainly and the preceding results are true practically.

To demonstrate the reconstruction capability of Zernike coefficients using the proposed sampling method, a surface made up of the forward 36-term Fringe Zernike polynomials whose coefficients are all 1 nm are constructed. The constructed surface is shown in Fig. 3(a). Then the Zernike coefficients are reconstructed using Eq. (15) with the perturbation

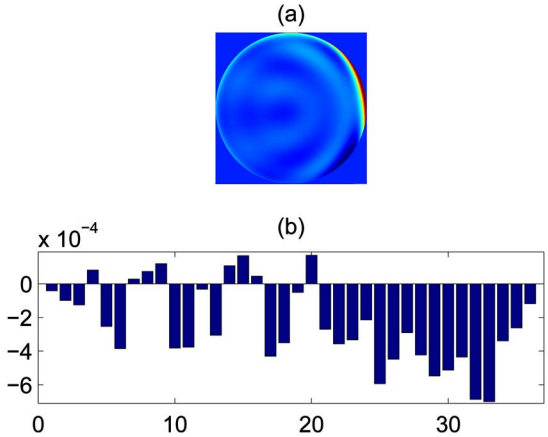

Fig. 3. (Color online) Zernike surface and reconstruction errors. (a) Constructed Zernike surface, and (b) differences of reconstructed and original coefficients.

points used in Fig. 2(d). The differences of the reconstructed and original 36-term Fringe Zernike coefficients are shown in Fig. 3(b).

One can see that the errors of Zernike reconstruction coefficients are of the 10^{-4} level and they are very small compared to the original coefficients that are all 1 nm. That is to say that the reconstruction relative errors are only about −4 orders of magnitude, which proves that the proposed sampling method for reconstruction of Zernike coefficients is very accurate. Compared to the conventional least squares fitting method, the proposed method is also more effective because only a few multiplication and addition operations are needed. Besides that, a linear equation group like $Ax = b$ is needed to solve the least squares fitting method. Usually the condition number of the matrix A is far more than 1; that is to say that the method is very sensitive to the perturbation of the measured or finite element analysis analyzed data.

5. Conclusion

In summary, a mathematically constructed sampling method that ensures the orthogonality of Zernike polynomials over the discrete points within the unit circle was introduced. And the constructed discrete points can assure the correctness of Zernike reconstruction coefficients. The proposed discretization meshes of the unit circle can be further exploited in optomechanical interface conversion software such as SigFit [6] or in wavefront measuring instruments like digital interferometers, where Zernike functions are commonly used to describe wavefronts or surface displacements.

References

1. G. Dai and V. N. Mahajan, "Orthonormal polynomials in wavefront analysis: error analysis," Appl. Opt. **47**, 3433–3445 (2008).
2. V. N. Mahajan, "Zernike polynomial and wavefront fitting," in *Optical Shop Testing*, D. Malacara, ed. (Wiley, 2007), pp. 498–546.
3. P. de Groot, "Design of error-compensating algorithms for sinusoidal phase shifting interferometry," Appl. Opt. **48**, 6788–6796 (2009).
4. Z. Shi, J. Zhang, Y. Sui, J. Peng, F. Yan, and H. Jiang, "Design of algorithms for phase shifting interferometry using self-convolution of the rectangle window," Opt. Express **19**, 14671–14681 (2011).
5. K. B. Doyle, V. L. Genberg, G. J. Michels, and G. R. Bisson, "Optical modeling of finite element surface displacements using commercial software," Proc. SPIE **5867**, 58670I (2005).
6. SigFit Reference Manual, Sigmadyne Inc. (2011).
7. J. C. Wyant and K. Creath, "Basic wavefront aberration theory for optical metrology," in *Applied Optics and Optical Engineering*, R. R. Shannon and J. C. Wyant, eds. (Academic, 1972), pp. 1–53.
8. M. Pap and F. Schipp, "Discrete orthogonality of Zernike functions," Mathematica Pannonica **16**, 137–144 (2005).