2024

Optical Differentiation Wavefront Sensor with Pixelated Filters Toward High Performance Metrology

Biswa R. Swain
bs3858@rit.edu

Optical Differentiation Wavefront Sensor with Pixelated Filters Toward High Performance Metrology

by

Biswa R. Swain

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Imaging Science
in the College of Science

Rochester Institute of Technology
Rochester, NY
2024

Doctoral Committee:

Professor Jie Qiao
Professor Grover Swartzlander
Professor Christopher Kanan
Professor Michael Zemcov

Table of Contents

Acknowledgements

I wish to express gratitude to my Advisor Professor Jie Qiao for providing outstanding support and guidance in my PhD journey. Under her guidance, I acquired both scientific and non-technical skills. I would like to acknowledge the National Science Foundation, the Center for Freeform Optics (CeFO), and the Rochester Institute of Technology for sponsoring this research. I also thank my committee members, Professor Grover Swartzlander, Professor Michael Zemcov and Professor Christopher Kanan for their inputs. I would like to thank Dr. Christophe Dorrer and Prof. Jannick Rolland at the University of Rochester for the research collaboration.

I thank Dr. Feng Tao for training me in experiments at the beginning stage of my PhD. I thank Scott DeFisher at OptiPro Systems for providing the reference measurements for some important parts of my work. I appreciate the assistance I received from Akif Qadeer and Renuka Manjula in the CeFO project.

I am thankful to my peers and labmates for offering discussions and words of encouragement. Finally, I wish to thank my parents, brother, friends and relatives for their love and support.

Abstract

Freeform optics enables enhanced system performance and packaging due to the high degrees of design freedom they provide, but characterizing these optics remains challenging. Phase measurement of an optical wave is essential in the metrology of optical components, adaptive optics, and laser beam quality assessment. The Shack-Hartman Wavefront Sensors are limited by their dynamic range and resolution due to lenslet operation. Interferometric-method-based adaptive nulling requires high-performance wavefront sensor, and coordinate-scanning-based measurements are time consuming. We investigate and develop an Optical Differentiation Wavefront Sensor (ODWS) based on binary pixelated filter that enables high dynamic range, high resolution freeform metrology. Analysis of experimental results and comparison with commercial metrology show that phase plates with different magnitude of wavefront slopes can be accurately characterized. We created an ODWS design that reduced the footprint by five times compared to typical $4f$ arrangement for the same effective focal length. We further investigated the system alignment tolerance. We report on the theoretical and experimental demonstration of an ODWS based on binary pixelated linear and nonlinear amplitude filtering in the far-field. We trained and tested a convolutional neural network that reconstructs spatial phase map from nonlinear filter based ODWS images where an analytic solution is not available. It shows accurate retrieval (~0.05λ root mean square error) over different magnitude of wavefronts and on random shaped wavefronts. This work paves the way for realizing simultaneous sensitive, high dynamic range and high-resolution wavefront sensing.

List of Tables

List of Figures

Chapter 1 Introduction

1.1 Overview of metrology techniques

Wavefront measurement is one of the key requirements in many areas of optics and, with applications such as metrology, astronomy, and laser engineering. The wavefront of an optical field is defined as the locus of points having the same phase. In metrology, the form of an optical component can be characterized by measuring the wavefront change it induces. Wavefront measurements in optical engineering allow for better fabrication and integration of optical components, and improvements in metrology are generally required to permit improvements in fabrication. This is particularly true for freeform optical components which simultaneously require high dynamic range and high-resolution measurements.

Freeform components are based on optical surfaces that have no rotational or translational symmetry which allows mitigating aberrations and compact packaging. Freeform optical components enable compact and high-performance imaging systems [1–3]. It finds applications in fields such as wide field-of-view telescopes, augmented and virtual reality, illumination, and medical technologies. Figure 1.1(a) shows a long wave infrared (LWIR) imaging system design with 3 freeform mirrors and from Fig. 1.1(b), it can achieve diffraction limited performance (0.01λ RMS wavefront error) over 10° field of view [4]. The manufacturing quality of freeform optics depends on the capabilities of the available metrology tools, but characterizing these optics remains challenging [5,6].

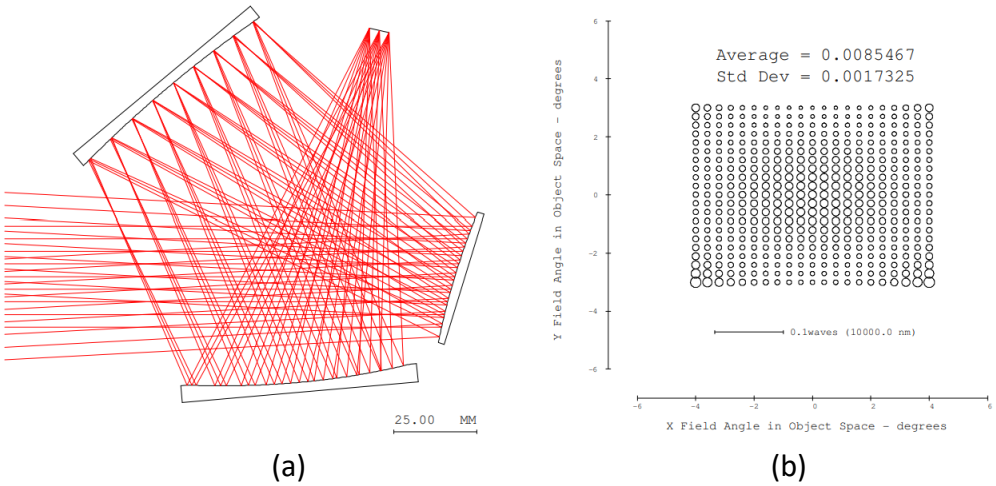

Figure 1.1 (a) Layout of an LWIR imaging system optimized with freeform surfaces and (b) the RMS wavefront error of the optimized system. Figure credit [4].

There are a multitude of techniques to obtain quantitative phase information [7]. Interferometry converts phase information into a fluence modulation that can be measured with a square-law photodetector. Interferometers that use a reference wave, e.g., Fizeau interferometers, are the workhorses of optical testing, but they are typically limited to operation close to null testing and weak aspheres. Typical interferometric nulling components such as computer-generated holograms are part-specific, alignment sensitive, and expensive. Adaptive interferometric null test using deformable mirror and Spatial Light Modulator (SLM) have been reported for freeform metrology [8,9]. Figure 1.2 shows the layout of a SLM based null test where the SLM serves as the adaptive nulling component to measure the freeform Surface Under Test (SUT) interferometrically [9]. Here the nulling wavefront generated by the SLM still needs to be precisely characterized with a high-performance wavefront sensor.

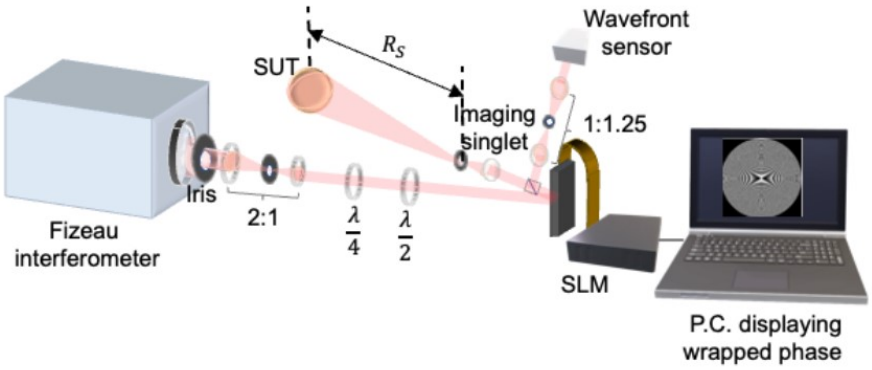

Figure 1.2 Schematic of a null test. The SLM serves as adaptive nulling component to measure the surface under test interferometrically. Figure credit [9].

Shack-Hartmann wavefront sensors (SHWS) measure the wavefront slopes obtained by determining the centroid displacement of far-field spots generated by an array of microlenses. Figure 1.3 illustrates this principle. The integration of the slopes produces the wavefront map.

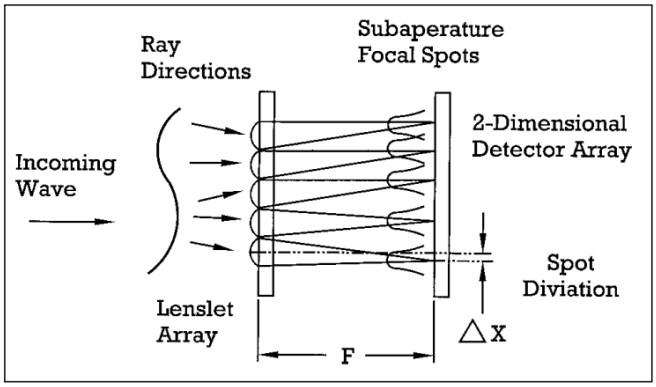

Figure 1.3 A Schematic showing the principle of SHWS. The shifts in the positions of the focused spots from each lens array determine the local wavefront tilts. Figure credit [10].

The inability to correctly identify the spots generated by the different microlenses limits the dynamic range of SHWS when these spots are significantly displaced by high wavefront slopes [11]. Dynamic range extension of SHWS based on the use of liquid crystals to switch the lenslets on and off has the disadvantages of light absorption, scattering and potential calibration issues [12,13]. A holographic SHWS that uses cross-correlation peak-displacement detection requires high-processing power and long processing time [14]. Sorting and phase unwrapping-based SHWS

methods are limited by maximum wavefront curvature [15,16], although not by the wavefront slope. Subaperture wavefront stitching with SHWS has been investigated to overcome the dynamic range issue [17, 18]. Other emerging techniques applied to freeform optical components include tilted wave interferometry [19] and low-coherence interferometry [20-22]. Although deflectometry can characterize complex shapes, it is prone to calibration errors [23, 24]. Optical profilometers based on the point-cloud method can measure large size optics with high slopes [25], but the measurement time increases with resolution and size of the optics. There is still significant gain in exploring and demonstrating new wavefront-sensing technologies because of the increasing role of freeform optics and the associated metrology challenges.

1.2 Motivation and Outline of the Dissertation

This project develops an alternative wavefront sensing methodology known as the Optical Differentiation Wavefront Sensor (ODWS). Sprague and Thompson showed that a large variation phase object can be visualized in an image having an irradiance that is directly proportional to the object phase [26]. Bortz pioneered the general implementation of an ODWS based on a rotating amplitude filter in a coherent processing setup [27]. In ODWS, the far-field of the test wave is modulated by an amplitude filter, and the resulting near-field fluence distribution is processed to yield the wavefront gradient in one direction. Two gradients in non-parallel directions, with orthogonal directions being a natural choice, are used to reconstruct the wavefront of the test wave [27-29]. The ODWS is potentially advantageous, in particular, in terms of adjustable high dynamic range, resolution, signal-to-noise ratio and achromaticity [28,29]. ODWS demonstrations in the literature based on different implementations of the amplitude filter mostly present simulations or the qualitative detection of phase variations [27-33].

Prior work synergistic with this project which was contributed by our team includes: creating and validating a $4f$-ODWS model, designing, and fabricating the amplitude filter, defining and evaluating parameters on ODWS performance, and measuring a rotationally symmetric lens [34].

Chapter 2 presents the theory of the ODWS principle and the wavefront reconstruction. The design of the amplitude filter based on binary pixelation is described and the impact of the pixel sizes are considered. The ODWS dynamic range and sensitivity are qualitatively discussed. Prior results on impact of focal length and pupil size are summarized. Chapter 3 presents the ODWS experimental setup, design baseline, and performance evaluation in freeform optics measurements for the first time to our knowledge. The effect of photodetection noise and non-ideal filter transmission profiles are studied. Chapter 4 details analytical derivation of a novel implementation of ODWS using telephoto lens for footprint reduction. Experimental results are presented, and alignment tolerance is investigated for instrumentation. Chapter 5 explores a Convolutional Neural Network (CNN) to reconstruct wavefronts from ODWS fluence measurements obtained with nonlinear filter transmission, potentially providing simultaneous high dynamic range and sensitivity. Simulation and experimental methods, and respective CNN training, testing results are presented. Chapter 6 presents simulation of a basic wavefront stitching algorithm. Finally, Chapter 7 summarizes our conclusion.

Chapter 2 Concept of Optical Differentiation Wavefront Sensor and Prior Work

2.1 ODWS concept and equation using 4f system

The concept of optical differentiation is based upon the idea that the first derivative of a function is obtained when its Fourier transform is multiplied by a linear function and an inverse Fourier transform of the product is taken. It is expressed as follows:

$$\frac{\partial u(x,y)}{\partial(x)} = \iint_{-\infty}^{\infty} j2\pi\xi U(\xi,\eta) e^{j2\pi(x\xi+y\eta)}\, d\xi d\eta, \tag{2.1}$$

where $u(x,y)$ and $U(\xi,\eta)$ are Fourier transform pairs. For determination of the optical phase $\varphi(x,y)$ in the near field, the ODWS relies on amplitude modulation in the far field of the optical wave under test, for example with a 4*f* optical system (Fig. 2.1).

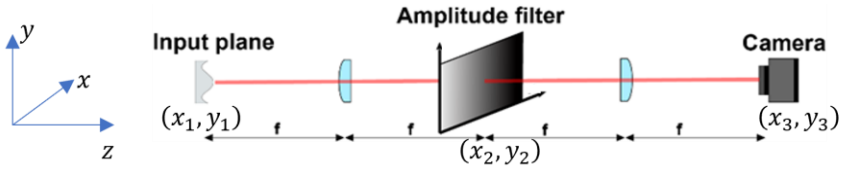

Figure 2.1 Principle of an ODWS, based on a 4f optical system.

Let $u_1(x_1, y_1)$ represent the field of the test object at the input plane. This field is Fourier transformed by the first lens to the far-field as

$$u_2(x_2,y_2) = \frac{1}{j\lambda f} U_1(\xi,\eta) = \frac{1}{j\lambda f} \iint_{-\infty}^{\infty} u_1(x_1,y_1) e^{\{j\frac{\pi}{\lambda f}[-2(x_1x_2+y_1y_2)]\}}\, dx_1 dy_1, \tag{2.2}$$

where (ξ,η) are the spatial frequencies defined as $\xi = \frac{x_2}{\lambda B}$ *and* $\eta = \frac{y_2}{\lambda B}$. $U_1(\xi,\eta)$ is the Fourier transform of input field. Note that the far-field filters can be linear or nonlinear. For the nonlinear filter, there is no closed form solution to retrieve the wavefront, and the deep learning based

solution is described in Chapter 5. We hereby describe the derivation for analytical solution for the former. The far-field is modulated by an amplitude of linear amplitude transmission (Eq. 2.3),

$$t = \frac{x_2}{W} + \frac{1}{2}, \tag{2.3}$$

W is the width of the filter. The second lens performs another Fourier transform of the modulated far-field to the detector plane, where the field can be written as:

$$u_3(x_3, y_3) = \frac{1}{(j\lambda f)} \iint_{-\infty}^{\infty} \left(\frac{x_2}{W} + \frac{1}{2}\right) \frac{1}{j\lambda f} U_1(\xi, \eta) e^{\{j\frac{\pi}{\lambda f}[-2(x_2 x_3 + y_2 y_3)]\}} \, dx_2 dy_2, \tag{2.4}$$

where the limits of the integration to ∞ is valid as long as the far-field $U_1(\xi, \eta)$ is contained within the filter width. Now using the differentiation property of Eq. (2.1), we can write Eq. (2.4) as:

$$u_3(x_3, y_3) = \frac{j\lambda f}{2\pi W} \left[\frac{\partial u_1(-x_3, -y_3)}{\partial(-x_3)} - \frac{1}{2} u_1(-x_3, -y_3)\right] \tag{2.5}$$

Now writing the input field in terms of its amplitude and phase as

$$u_1(x_1, y_1) = A(x_1, y_1) e^{j\varphi(x_1, y_1)}, \tag{2.6}$$

Eq. (2.5) can be written as,

$$u_3(x_3, y_3) = e^{j\varphi(-x_3, -y_3)} \left[\frac{j\lambda f}{2\pi W} \left(\frac{\partial A(-x_3, -y_3)}{\partial(-x_3)} + j \frac{\partial \varphi(-x_3, -y_3)}{\partial(-x_3)} A(-x_3, -y_3)\right) - \frac{1}{2} A(-x_3, -y_3)\right] \tag{2.7}$$

We define fluence as the optical energy delivered per unit area on the detector and obtained as the time integrated value of the optical intensity. The fluences are obtained by normalizing the measured intensity using the respective integration or exposure times (order of few milli-seconds) for different filter orientations. The detection fluence is then,

$$F_x(x_3, y_3) = u_3 u_3{}^* = \frac{\lambda^2 f^2}{4\pi^2 W^2} A^2(-x_3, -y_3) \left(\frac{\partial \varphi(-x_3, -y_3)}{\partial(-x_3)}\right)^2 + \frac{\lambda f}{2\pi W} A^2(-x_3, -y_3) \frac{\partial \varphi(-x_3, -y_3)}{\partial(-x_3)} +$$

$$\frac{\lambda^2 f^2}{4\pi^2 W^2} \left(\frac{\partial A(-x_3, -y_3)}{\partial(-x_3)}\right)^2 + \frac{1}{4} A^2(-x_3, -y_3) \tag{2.8}$$

If there was no far-field amplitude modulation, i.e., $t = 1$ in Eq. (2.3), then the corresponding fluence will be:

$$F_0(x_3, y_3) = A^2(-x_3, -y_3) \tag{2.9}$$

Now we make an approximation that the input field amplitude is constant i.e., $A(x_1, y_1) \approx A_0$ and that $\frac{\partial A(x_1,y_1)}{\partial(x_1)} = 0$. Using this approximation and Eq. (2.9) in Eq. (2.8), we obtain:

$$\frac{F_x(x_3,y_3)}{F_0(x_3,y_3)} = \left[\frac{\lambda f}{2\pi W}\frac{\partial\varphi(-x_3,-y_3)}{\partial(-x_3)} + \frac{1}{2}\right]^2, \tag{2.10}$$

From Eq. (2.10), the wavefront gradient along x-direction is thus,

$$\frac{\partial\varphi(-x_3,-y_3)}{\partial(-x_3)} = \frac{\pi W}{\lambda f}\left(2\sqrt{\frac{F_x}{F_0}} - 1\right) \tag{2.11}$$

The negative sign in coordinates simply means the spatial inversion of the field with respect to that at the input. The orthogonal phase slope is obtained after 90-degree rotation of the same filter to induce a transmission gradient oriented along *y*-direction, leading to the fluence distribution F_y and the slope,

$$\frac{\partial\varphi(-x_3,-y_3)}{\partial(-y_3)} = \frac{\pi W}{\lambda f}\left(2\sqrt{\frac{F_y}{F_0}} - 1\right) \tag{2.12}$$

Reconstruction of the phase φ from the phase gradients given by Eqs. (2.11) and (2.12) is equivalent to reconstructing the phase from SHWS data. The ODWS is advantageous because slope data are available at every pixel of the detected experimental trace, while slope data are intrinsically averaged over each lenslet in a SHWS.

2.2 Wavefront reconstruction algorithm

In this work, the wavefront is reconstructed from the two orthogonal wavefront slope measurements of dimensions $N \times N$ by applying the Southwell iterative procedure [35,36]. The Southwell algorithm uses a least square algorithm and thus reduces the impact of noise compared to direct local slope integration [36]. It uses global optimization to reconstruct the best fit wavefront compared to local slope integration. We briefly describe Southwell procedure below.

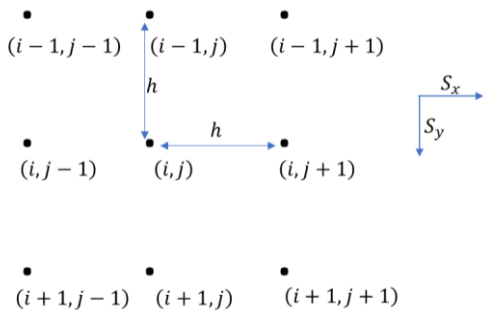

Figure 2.2 Nine adjacent grid points for the ODWS wavefront slope geometry. S_x and S_y represent horizontal and vertical slope vectors.

Figure 2.2 shows the nine grid points denoted by row index and a column index, ranging from $i-1$ to $i+1$ and $j-1$ to $j+1$, respectively. This indexing spans from the top left corner to the bottom right corner. In the Southwell algorithm, the average slope either horizontally or vertically at an intermediary point between two neighboring grid points in the respective direction is described in relation to the disparity in wavefront phase at the two adjacent grid points in the same direction. If the phase, horizontal slope, and vertical slope at the position (i,j) are represented as $\varphi^{i,j}, {S_x}^{i,j}, {S_y}^{i,j}$ respectively, at the position (i,j), the Southwell algorithm can be expressed as follows:

$$\frac{{S_x}^{i,j-1}+{S_x}^{i,j}}{2}=\frac{\varphi^{i,j}-\varphi^{i,j-1}}{h}$$

$$\frac{{S_x}^{i,j+1}+{S_x}^{i,j}}{2}=\frac{\varphi^{i,j+1}-\varphi^{i,j}}{h}$$

$$\frac{{S_y}^{i-1,j}+{S_y}^{i,j}}{2}=\frac{\varphi^{i,j}-\varphi^{i-1,j}}{h}$$

$$\frac{{S_y}^{i+1,j}+{S_y}^{i,j}}{2}=\frac{\varphi^{i+1,j}-\varphi^{i,j}}{h} \tag{2.13}$$

where h is the distance between two adjacent grid points in horizontal or vertical direction. Upon solving Eq. (2.13) we can obtain:

$$\varphi^{i,j} = \frac{\varphi^{i-1,j}+\varphi^{i,j-1}+\varphi^{i+1,j}+\varphi^{i,j+1}}{4} + \frac{h}{8}\left[\left(S_x{}^{i,j-1} + S_x{}^{i,j}\right) - \left(S_x{}^{i,j+1} + S_x{}^{i,j}\right) + \left(S_y{}^{i-1,j} + S_y{}^{i,j}\right) - \left(S_y{}^{i-1,j} + S_y{}^{i,j}\right)\right] \quad (2.14)$$

Equation (2.14) can be rewritten in general form as:

$$\varphi^{i,j} = \overline{\varphi^{i,j}} + b^{i,j}/g^{i,j} \quad (2.15)$$

where $\overline{\varphi^{i,j}}$ is the mean nearest neighbor phase, $b^{i,j}$ is a constant depending on the slope measurements, and $g^{i,j}$ is either 4, 3 or 2 depending on whether the phase point $\varphi^{i,j}$ is an interior point, an edge, or a corner respectively. Equation (2.15) is solved iteratively. The calculation of the right-hand side of Eq. (2.15) involves the utilization of current values for the φ's, which may be initiated as zeros. The resultant output then serves as an improved estimation. At the m^{th} iteration, this is represented as:

$$\varphi^{i,j}{}_{(m+1)} = \overline{\varphi^{i,j}}_{(m)} + b^{i,j}/g^{i,j} \quad (2.16)$$

The successive over-relaxation (SOR) iterative technique is used in which a phase point is updated based on the value of the previous iteration at that point. Equation (2.16) can be modified as:

$$\varphi^{i,j}{}_{(m+1)} = \varphi^{i,j}{}_{(m)} + \omega\left[\overline{\varphi^{i,j}}_{(m)} + b^{i,j}/g^{i,j} - \varphi^{i,j}{}_{(m)}\right] \quad (2.17)$$

where ω is a relaxation parameter which optimal value is given as [36]:

$$\omega = \frac{2}{1+\sin[\pi/(N+1)]} \quad (2.18)$$

We note that any algorithm developed for wavefront reconstruction from wavefront slopes, e.g., for Shack-Hartmann wavefront sensors, can be used, including modal reconstruction. One example of modal reconstruction is wavefront decomposition on an orthogonal basis of Zernike polynomials. The Southwell procedure was found to give excellent reconstruction on simulated and experimental data, below 0.05λ (λ = 633nm) RMS error [34]. It can operate on arbitrarily

shaped apertures and does not require the wavefront to be expandable on a particular basis, which allows for application to the widest range of wavefront metrology applications (note for example that the freeform phase plates used in the experimental demonstration (see Chapter-3) have a manufacturing artifact that is poorly fitted by the low-order modes).

2.3 Filter implementations

Despite the potential advantages of ODWS, it is not widely used mainly because of difficulty in practical realization of well-controlled transmission filter. Already published ODWS demonstrations include the use of a liquid crystal SLM [37-39], spatially varying optical activity [40,41] and holographic film [42]. A programmable amplitude filter can be created by placing an SLM between two polarizers. The transmission will vary based on the pixel signal applied and the wavelength due to the wavelength-dependent phase retardation. SLM-based filters are both wavelength and polarization dependent and may have undesirable diffraction orders. Spatially varying optical activity method such as placing polarization rotator between two Wollaston prisms requires fabrication of a custom optical component and is wavelength-sensitive. Holographic film might not have the optical quality required for accurate wavefront reconstruction.

2.4 Binary pixelated filter concept

In the ODWS configuration (Fig. 2.1), consider $E(x, y)$ to be the input field, and $s(u, v)$ to be the transmission of binary pixelated mask where s is either 0 or 1. The electric field just after the mask is $\tilde{E}(u, v) \times s(u, v)$, where $\tilde{E}$ is the Fourier transform of the input field E. The resulting field in the image plane is then another Fourier transform of the modulated field which can be written using convolution theorem as $E(x, y) \otimes \tilde{s}(x, y)$, which is convolution (denoted as $\otimes$) of the input field with the Fourier transform of the binary mask. If t is the ideal continuous transmission mask,

then designing of the pixelated mask requires minimizing the error $E \otimes (\tilde{s} - \tilde{t})$ over the output aperture. Now minimizing the error $(\tilde{s} - \tilde{t})$ in RMS sense is equivalent to the case of minimizing the transmission error $(s - t)$ when the filter is in near field configuration [43]. The latter case of filter in near field is a beam shaping configuration for which the binary pixelated filter design has been studied [43,44].

We have realized the binary pixelated transmission mask using the error diffusion algorithm [44]. Figure 2.3 illustrates the method. Error diffusion follows a sequential processing of mask pixels, typically from top to bottom and left to right. Fig. 2.3(a) demonstrates the representation of the target transmission of the shaper, $t(m, n)$, while Fig. 2.3(b) depicts the design of the binary shaper. In the latter, processed pixels with transmission $s(m, n)$ are binary (0 or 1), displayed in black or white, whereas unprocessed pixels are shown in gray. As the target transmission spans from 0 to 1 and the shaper transmission is binary, assigning each pixel value in the binary mask induces a transmission error.

In the error diffusion process, the binary value of $s(m, n)$ depends on the target shaper transmission $t(m, n)$ and the values of the transmission error at previously processed pixels. Initially, $t(m, n)$ is initialized as the square root of $I(m, n)$, where I is the sampled representation of the target intensity of the coherent beam. Referring to the flowchart in Fig. 2.3(c), the assignment of the transmission, $s(m, n)$, is determined by comparing the target shaper transmission to 0.5. If the target transmission is less than 0.5, $s(m, n)$ is set to 0, and if the transmission is greater than 0.5, $s(m, n)$ is set to 1.

The resulting error is spread to unprocessed neighboring pixels to bias the binary choice for these pixels and locally compensate for the transmission error. This is achieved by adding a fraction of the error $e(m, n) = s(m, n) - t(m, n)$ to the target transmission for these pixels, as

indicated by white arrows in Fig. 2.4(a). The target transmission, $t(m+a, n+b)$, is then updated to $t(m+a, n+b) + c(a,b) \times e(m,n)$ for the chosen set of integers a and b. Here only the four neighboring pixels at coordinates (m +1, n −1), (m+1, n), (m +1, n +1) and (m, n+1) are used in the four-weight error diffusion process. Function c is defined with values $c\,(1,-1) = -3/16$, $c(1,0) = -5/16, c(1,1) = -1/16$, and $c(0,1) = -7/16$. The algorithm then continues with the subsequent pixel as per the lexicographical order as per Fig. 2.3(b).

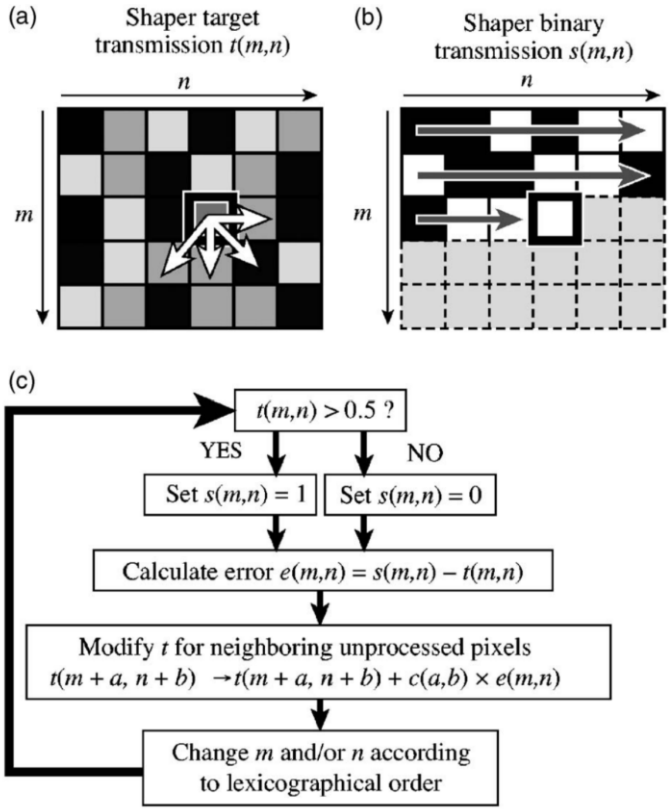

Figure 2.3 The error diffusion algorithm, (a) Example target beam shaper transmission; (b) Corresponding binary shaper in design; (c) Flowchart of the design process. Figure credit [44].

In the recent work of our Group [34], we have used spatially dithered distributions of transparent and opaque pixels to synthesize the transmission profile of an ODWS filter. Figure 2.4 shows a design example of a binary pixelated linear amplitude filter. These components are

achromatic and can be fabricated at large aperture and low cost by commercial lithography of a metal layer deposited on a glass substrate.

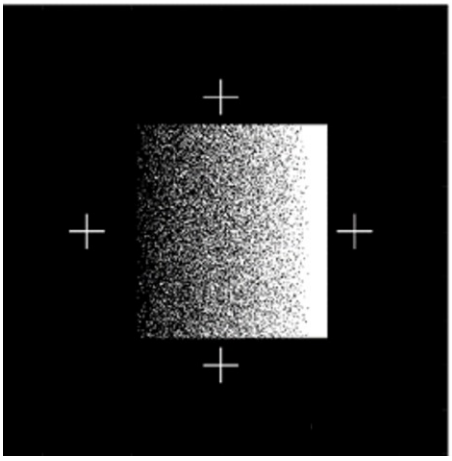

Figure 2.4 Design example of binary pixelated linear amplitude filter.

2.5 Dynamic range and sensitivity

Dynamic range is defined as the range of the wavefront slopes that can be measured. The ODWS solution in the Section 2.1 requires the far-field of the input field to be contained in the filter which ranges from $-W/2$ to $W/2$. From Eqs. (2.11) and (2.12), the dynamic range is, DR = $2\pi W/(\lambda f)$. Thus the DR is directly proportional to the filter width and inversely proportional to the focal length for a given wavelength. Implementation of larger filter width can be enabled by the binary pixelated transmission using metal-on-glass lithography technique. Decreasing the focal length is feasible in the case of ideal continuous filter, but for binary pixelated filter the wavefront retrieval accuracy depends on ratio of the far-field size and pixel size and is thus limited.

From Eqs. (2.11) and (2.12), the difference in fluence images, with and without the filter ($F_{x,}$ *or* $F_{y,}$ and F_0), gets reduced with increasing filter width for a given wavefront slope. Thus, when detection noise is present, the measurement of small phase variations will be corrupted by a large filter width, leading to the decrease in sensitivity. Therefore, there is a tradeoff between

dynamic range and sensitivity. This trade off can be reduced by non-linear filter transmission profile, providing advantage over SHWS.

2.6 Filter pixel size considerations

The spatial dithering algorithm based pixelated filter creates noise that is concentrated at higher frequencies and lacks density at the zero spatial frequency [44]. Since a binary pixelated filter is employed in the far field of the ODWS, the noise from the filter becomes visible in the detection plane. This leads to a decrease in the accuracy of the ODWS compared to an ideal ODWS that has a continuous filter, as the filter noise disrupts the precise measurement of fluences $F_{x,}$ and F_y.

Performance simulations of binary pixelated filter ODWS were done with test wavefronts and were compared with that of an ideal filter ODWS [34]. The system consists of a $4f$ line with using lenses with a focal length $f = 1$ meter, $\lambda = 633$nm, the filter width is set at 1 cm ($W = 1$cm), and a 1-cm pupil diameter. The dynamic range is 100 rad/mm. Figure 2.5 illustrates the simulated fluence F_y in the detection plane using a logarithmic scale for a flat input wavefront. It displays data for both a continuous filter and binary pixelated filters with pixel sizes of 10 μm, 5 μm, and 2.5 μm. The presence of noise generated from the binary pixels is evident. The high-frequency noise from pixelation and binarization is pushed away from the optical axis as the pixel size decreases, i.e., away from the camera's field of view, resulting in accurate implementation of the amplitude filter.

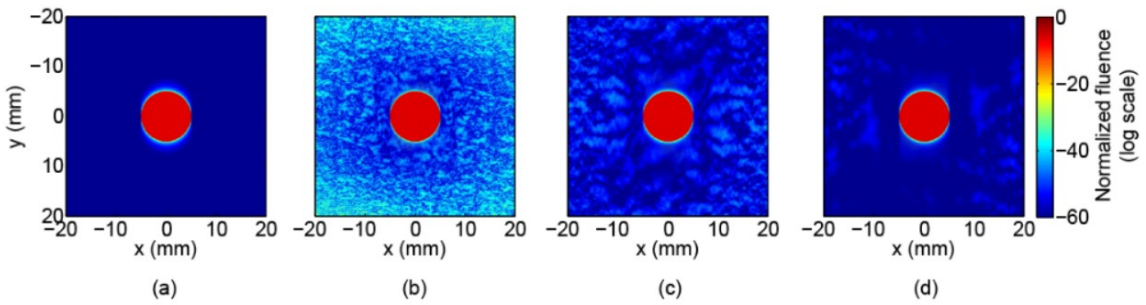

Figure 2.5 Fluence in image plane plotted on a logarithmic scale for (a) continuous filter; (b), (c), and (d) binary pixelated filters with pixel sizes 10 μm, 5 μm, and 2.5 μm respectively. Figure credit [34].

The smallest far-field size for a given input pupil size S is approximately $\lambda f/S$. Consequently, to ensure performance comparable to that of an ideal ODWS, it is necessary for the pixel size to be much smaller than $\lambda f/S$ for a binary pixelated filter. Performance for input wavefront of profile $\varphi(x,y) = \varphi_0 \cos(2\pi\,(x+y)/p)$ was simulated which has the slopes in the limit $[-2\pi\varphi_0/p\,, 2\pi\varphi_0/p]$ [34]. For $p = 1$ mm a decrease in performance is anticipated due to finite dynamic range when φ_0 reaches approximately 8 radians. The findings are presented in Figs. 2.6(a) and 2.6(b). The ideal ODWS exhibits excellent performance, with RMS and peak errors below λ/100 across the modulation amplitude range from 0 to 2π. The errors for different binary pixelation variations remain relatively consistent over the φ_0 range of 0 to 2π. The error appears to be directly proportional to the size of the pixel. The performance decrease at φ_0=8 rad is observed due to the dynamic range limit.

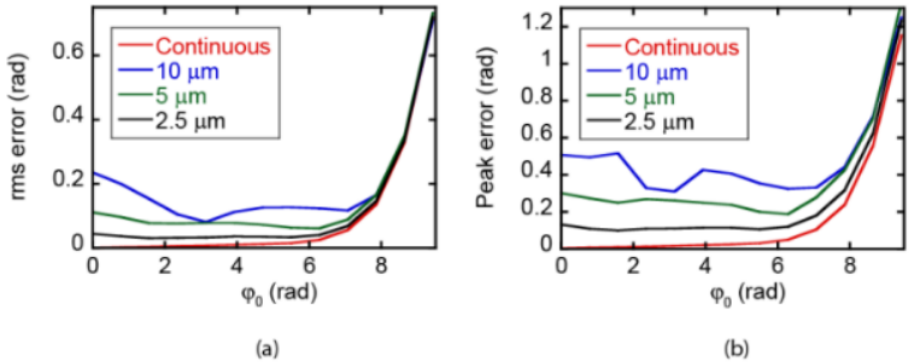

Figure 2.6 (a) RMS and (b) peak errors vs. the amplitude of phase modulation for continuous and different pixelated filters. Figure credit [34].

2.7 Effect of pupil size and focal length

Enlarging the pupil size results in the reduction in the far field feature size for a given focal length. This reduction in detail impacts the accuracy of measurements when utilizing a binary pixelated filter of a given pixel size. Figure 2.7(a) shows the RMS error in relation to the amplitude of a sinusoidal wave with a 1-mm period across a 2-cm pupil. Decrease in accuracy of about a factor of two is observed for the 2-cm pupil compared to the case of 1-cm pupil (in Fig. 2.6(a)). The error increases sharply as φ_0 approaches 8 radians due to the dynamic range limit which is independent of the pupil size. For measuring larger pupils, the subaperture stitching method can be employed and is discussed in Chapter-6.

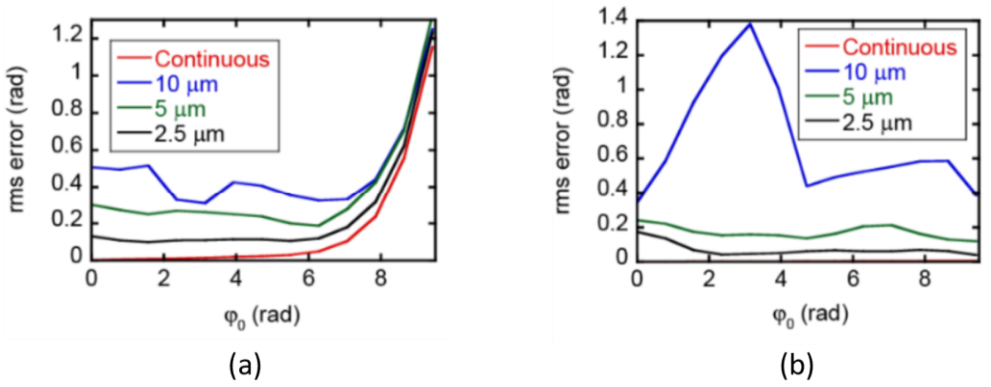

Figure 2.7 RMS errors vs. the amplitude of phase modulation for continuous and different pixelated filters (a) for a 2-cm-diameter pupil and f=1m, (b) for a 1-cm-diameter pupil and f=0.5m. Figure credit [34].

The dynamic range of ODWS varies inversely with the focal length, and a decrease in the focal length permits the far-field filter to accommodate a broader range of wavefront slopes. However, the drawback of a shorter focal length is the reduction in the scale of the features in the far field, leading to decreased accuracy in reconstruction. Figures 2.7(b) illustrates this consequence for a halving of the focal length for the 1-mm-period sinewave wavefront profiles. Beyond the amplitude value φ_0 = 8 radians, no decline in accuracy is observed, indicating a larger

dynamic range. However, the reconstruction errors are approximately twice than those of the 1-meter lens setup (in Fig. 2.6(a)).

Chapter 3 Demonstration of an ODWS in a 4f configuration for freeform metrology

This chapter presents the experimental setup, design baseline, and performance evaluation in freeform phase plate measurements. We also evaluate the effect of photodetection noise and non-ideal filter transmission profiles in simulation.

3.1 Integrate an in-situ testbed

We integrated a setup as shown in Fig. 3.1. A 20-mm collimated laser beam (λ = 633 nm) propagates through a test phase plate. The image-relay system (focal lengths of 750 mm and 125 mm) down-collimates the beam by a factor of 6.1. The field is Fourier transformed by the first lens of a 4*f* system (*f* = 1 m) to a Fourier plane where the gradient transmission filter is located. After amplitude modulation, the field is Fourier transformed back by the second lens and the resulting fluence is finally detected by a CCD camera. It is noted that, down collimation does not change the Peak-To-Valley (PV), nor Root Mean Square (RMS) of the wavefront. However, the wavefront slope increases proportionally to the down-collimation factor. The spatial resolution of the ODWS measurement at the phase plate location is 26.7 μm, corresponding to the camera pixel resolution of 4.4 μm. The camera has an 8-bit analog-to-digital converter.

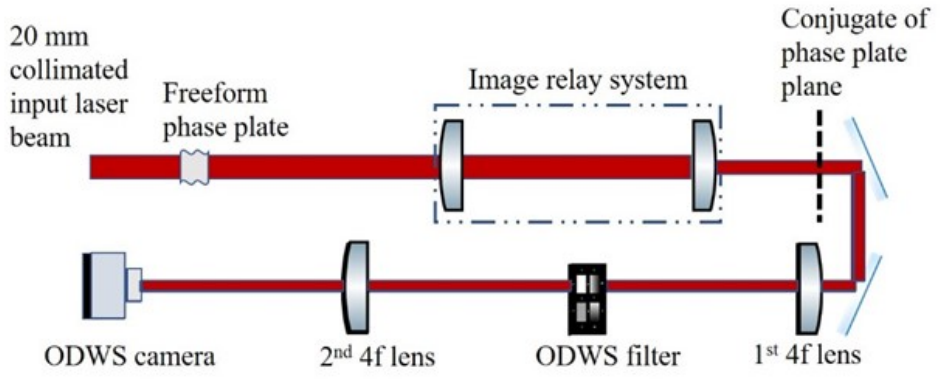

Figure 3.1 ODWS system layout.

Spatially dithered amplitude filters with field transmission changing linearly between 0 and 1 over the width W = 10 mm were designed using the standard four-weight error diffusion algorithm (see Section 2.4). Although filters with pixels size ranging from 2.5 μm to 10 μm have been manufactured and tested, the experimental results presented in this work were obtained with the 2.5-μm filter because the ODWS accuracy increases with decreasing pixel size of binary filter (Fig. 2.6). Although the 2.5-μm filter has the highest nonlinearity relative to the ideal ODWS-filter profile (see Section 3.5), it was experimentally found that it yields better retrieval in practice because of lower pixelation noise. The slope dynamic range at ODWS object plane is ~16 waves/mm, corresponding to ~20 μm/mm surface slope at 633-nm wavelength. The bias point in our experiment is the center of the filter, hence the range of wavefront slopes that can be measured at the conjugate of phase plate is [-8, +8] waves/mm. The corresponding dynamic range at the phase plate location is [-2.3, +2.3] waves/mm.

3.2 Establish reference measurements for the phase plates to be tested

To assess the performance of the ODWS, we have characterized one rotationally symmetric phase plate and three freeform phase plates. The freeform phase plates (thereafter referenced to as #1, #2, and #3) were designed to have horizontal coma of different amplitude over a 20-mm aperture

to investigate the ODWS performance over a range of wavefront slopes. Due to manufacturing error, they contain power and other aberrations. The corresponding range of wavefront slopes at the conjugate plane of phase plates in waves/mm are [-3.4, 1.9], [-11.1, 3.7] and [-5.5, 19.3]. The wavefront slopes of the first two phase plates are either within or slightly exceeding the ODWS design baseline and that of third phase plate is higher than the design baseline.

The four phase plates were fabricated with UltraForm [45] at OptiPro Systems. These parts were characterized by an UltraSurf (at OptiPro Systems) metrology system [46]. Figure 3.2(a) shows an UltraSurf system. It is a computer-controlled, non-contact coordinate measuring machine equipped with five air-bearing axes, linear motors, high-resolution feedback, and a non-contact probe. The measuring probe scans over the optical surface normal to it. This perpendicularity is maintained by adhering to the programmed nominal shape into the machine. UltraSurf can rotate the sensor ±130° from the vertical to allow measurement of steep slopes. High-resolution scanning allows for the examination of surface texture in small areas. The ability to measure large areas with steep slopes and high departure makes it possible to characterize various freeform shapes. The measurement generates a three dimensional point cloud which is compared against the nominal shape. The probe used for the thickness measurements of the phase plates in this work uses low-coherence interferometry (LCI) to gauge the thickness of multiple layers at a singular point, shown in Fig. 3.2(b).

Figure 3.3(a) shows the typical arrangement for low-coherence interferometry [47]. The light from a low-coherence source is split into the sample arm and the reference arm through the fiber coupler. The light reflected from both the sample and reference arms is then recombined in the fiber coupler and is redirected toward the detector. Optical interference occurs when the optical path lengths of the beams reflected by the sample reflector and the reference mirror differ by less

than the finite coherence length of the source. When measuring a single reflector, the reference mirror is scanned. Figure 3.3(b) shows the typical signal obtained which constitutes constructive and destructive interferences. The pattern is modulated by an envelope related to the coherence function of the source. If the sample has spaced reflectors (like phase plates), a specific envelope function is registered for each reflector. The precise position of each reflector can be identified by the exact location of the maximum point on the envelope function.

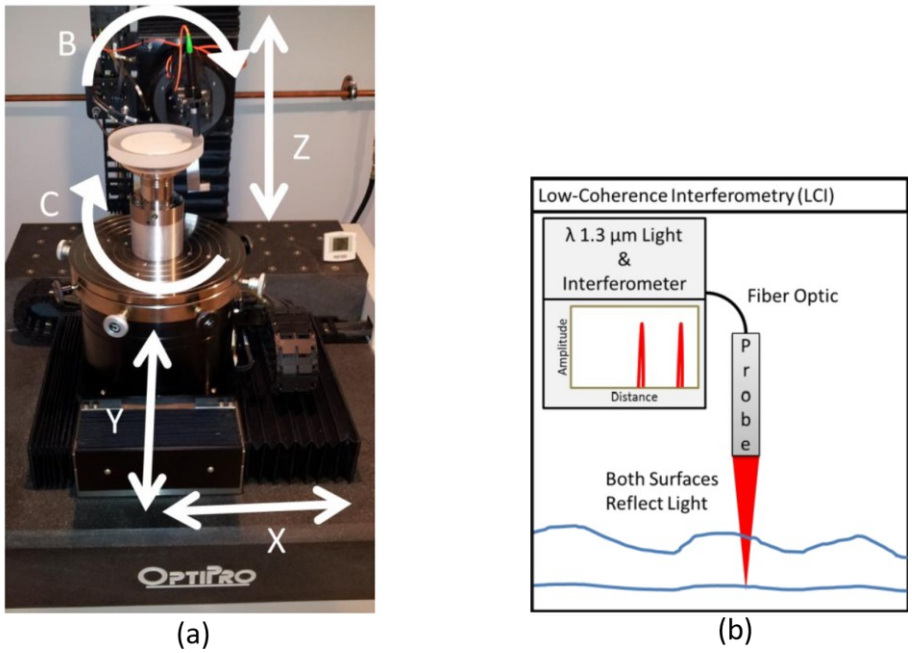

Figure 3.2 (a) UltraSurf, featuring labeled axes, has the probe attached to the B-Z axes, while the part is mounted to the X-Y-C axes, (b) Operation of the LCI probe. Figure credit [46].

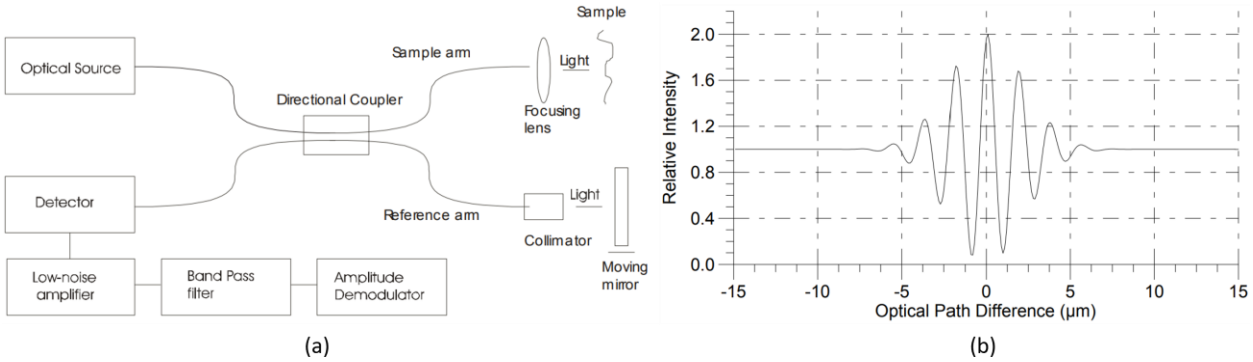

Figure 3.3 (a) Typical LCI configuration with fiber-based system, (b) Interference signal from a single reflector. Figure credit [47].

Characterizing the accuracy of measured figure error on UltraSurf is challenging in a generalized manner due to the multi-axis nature of the system. Different geometries will have different levels of uncertainty. UltraSurf qualified flat and sphere to 0.1λ (He-Ne) peak to valley with an interferometer [46]. Therefore, we use 0.1λ RMS as benchmark to qualify ODWS measurements with UltraSurf in this work. The UltraSurf measurements were obtained at a spatial resolution of 250 μm; the measurements were interpolated for direct comparison with ODWS.

Characterizing the freeform phase plates with a Zygo interferometer (Zygo GPI 4" Verifier AT) in transmission mode without any nulling component was successful for phase plate #1 but was not successful for phase plates #2 and #3, which would therefore require the manufacturing of specific nulling components. We therefore used the residual errors between ODWS and UltraSurf measurements to validate the ODWS measurements. It should be noted that this error contains the inherent error in both the ODWS and UltraSurf measurements, as well as imperfect registration (translation and rotation) between the reconstructed wavefronts, magnification and other experimental imperfection. For all wavefronts, piston, tip, and tilt were removed, as is customary in the metrology of optical components.

3.3 Phase plate measurement results

The ODWS was first compared to the UltraSurf when characterizing a rotationally symmetric optics. The UltraSurf measurement is shown in Fig. 3.4(a). Figure 3.4(b) shows the wavefront result as an average of 20 sets of measurements by the ODWS. The validity is determined by calculating the RMS of difference between the measured wavefronts. The difference is shown in Fig. 3.4(c) and the measurements agree to 0.08λ.

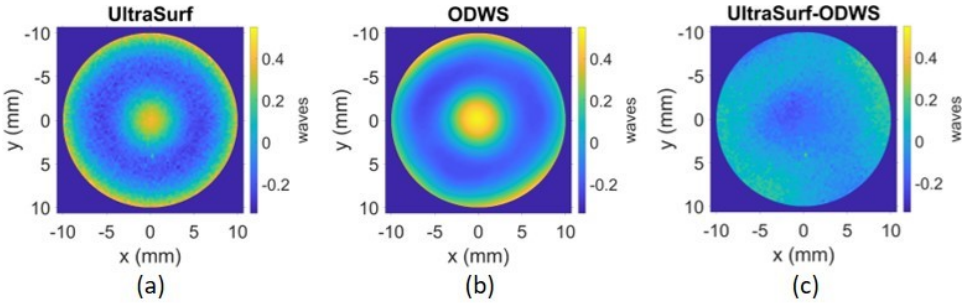

Figure 3.4 Wavefront of a rotationally symmetric optics (a) measured with UltraSurf, RMS = 0.18λ, PV = 0.8λ; (b) measured with ODWS, RMS = 0.19λ, PV = 0.8λ (c) difference, RMS = 0.08λ.

We present and compare the transmission wavefront measured by the ODWS, the UltraSurf and the Zygo interferometer for freeform phase plates #1 and #2. The wavefront of phase plate #1 measured by UltraSurf and ODWS are shown in Fig. 3.5(a) and 3.5(b) respectively, and the difference of the two measurements is shown in Fig. 3.5(c). A similar comparison between Zygo interferometer and ODWS measurement for phase plate #1 is shown in Fig. 3.6. Similarly, the measurement results of phase plate #2 are shown in Fig. 3.7. The ODWS wavefront result is obtained by averaging the measurements performed on four different days, each having 10 measurements. Piston, tip/tilt were removed due to the reasons mentioned in Section 3.2. The feature located at the center of each phase plate, which is retrieved by the three diagnostics, is a manufacturing artifact. From the difference maps shown in Figs. 3.5(c), 3.6(c) and 3.7(c), the wavefront measured on phase plates #1 and #2 by the ODWS are comparable to the wavefront measured by other techniques. The corresponding RMS of the wavefront difference is 0.04λ (phase plate #1, ODWS/UltraSurf), 0.06λ (phase plate #1, ODWS/Zygo) and 0.1λ (phase plate #2, ODWS/UltraSurf). For reference, the RMS of the wavefront difference calculated with the Zygo and UltraSurf wavefront maps for phase plate #1 is 0.05λ, showing that even the wavefront returned by the two commercial diagnostics are not identical.

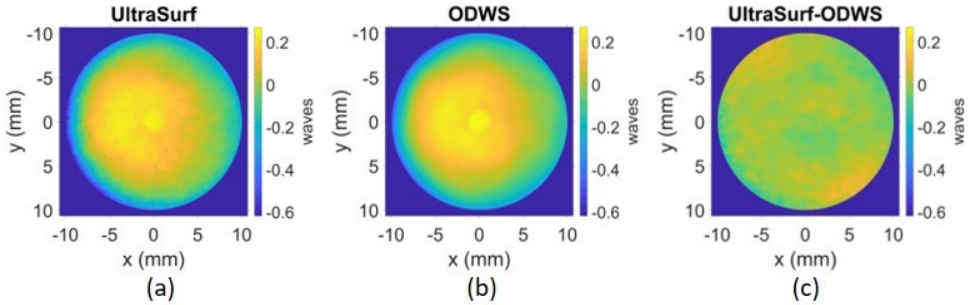

Figure 3.5 Wavefront of freeform phase plate #1 (a) measured with UltraSurf, RMS = 0.16λ, PV = 0.86λ; (b) measured with ODWS, RMS = 0.15λ, PV = 0.79λ; (c) difference, RMS = 0.04λ.

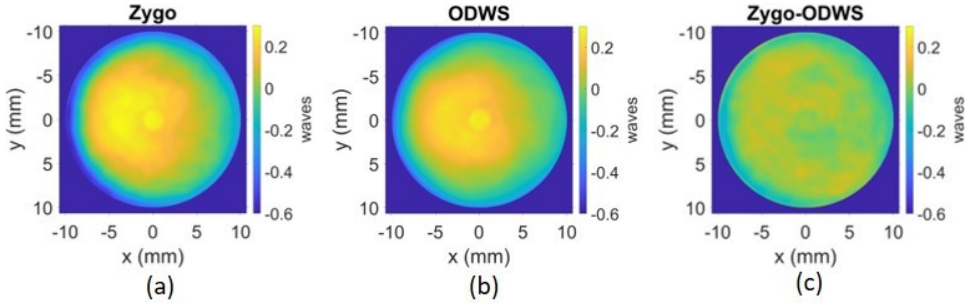

Figure 3.6 Wavefront of freeform phase plate #1 (a) measured with Zygo interferometer, RMS = 0.2λ, PV = 0.87λ; (b) measured with ODWS, RMS = 0.15λ, PV = 0.79λ; (c) difference RMS = 0.06λ.

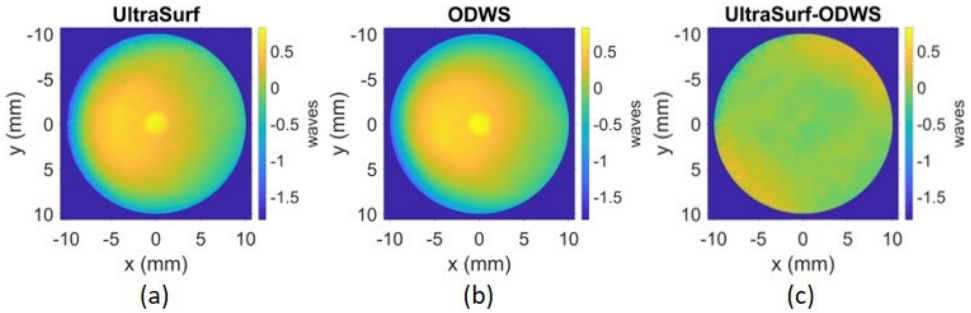

Figure 3.7 Wavefront of freeform phase plate #2 (a) measured with UltraSurf, RMS = 0.38λ, PV = 2.58λ; (b) measured with ODWS, RMS = 0.39λ, PV = 2.28λ; (c) difference RMS = 0.1λ

To illustrate the measurement for the freeform component of the wavefront, we removed the power term from the ODWS measurements shown in Figs. 3.5(b) and 3.7(b) and show the results in Figs. 3.8(a) and 3.8(b) for phase plate #1 and #2, respectively. The primary coma aberration is more clearly visible after the power removal, although the plotted wavefronts still include higher-order aberrations and the sharp central feature.

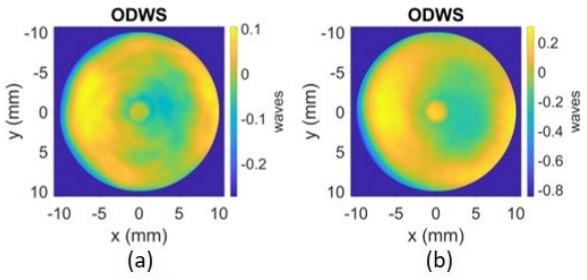

Figure 3.8 Wavefront measured by the ODWS after power removal for (a) freeform phase plate #1; RMS = 0.06λ, PV = 0.38λ; (b) freeform phase plate #2; RMS = 0.16λ, PV = 1.13λ.

The measurement precision was quantified by taking the RMS difference between individual wavefront maps and the average of multiple measurements on each day. Figure 3.9 shows an average precision of 0.014λ for phase plate #2, using data taken over four consecutive days. The maximum variation between measurements is 0.02λ at worst which shows that the measurements are highly consistent. The observed precision is impacted not only by detection noise but also by environmental factors such as air turbulence in the laboratory environment. Because this ODWS implementation is based on fluence measurements that are sequentially measured, the measurement precision could be improved using a stabilized laser source and a high-dynamic-range camera.

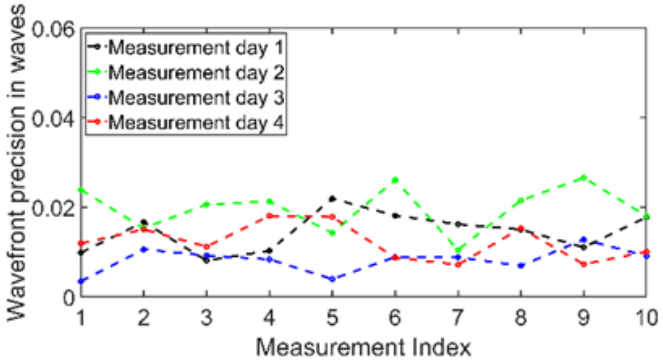

Figure 3.9 Precision plot for 40 measurements of phase plate #2 taken over four days, average precision is 0.014λ.

3.4 Impact of filter width on high wavefront slope measurement

High wavefront slopes leading to far-field components beyond the 1-cm filter width are not expected to be accurately measured, leading to poor wavefront reconstruction at the corresponding

near-field locations. The purpose of this section is to experimentally investigate this effect. For this study, we chose phase plate #3 which has wavefront slopes much higher than the theoretical limitation of the current ODWS design. The maximum wavefront slope resides near the left edge of the phase plate with an aperture of 20 mm. Because the portion of the beam that has larger wavefront slope lands closer to the edge of the filter, we compare wavefront measurements over two apertures with diameter equal to 15 mm and 20 mm, respectively. The ranges of wavefront slopes at the conjugate plane are within ([-5.5, 6.3] waves/mm for the 15-mm diameter) and much beyond ([-5.5, 19.3] waves/mm for the 20-mm diameter) the dynamic range design baseline for the implemented ODWS ([-8, 8] waves/mm, corresponding to the 1-cm filter width).

The wavefront maps of phase plate #3 in the 15 mm pupil are shown in Figs. 3.10(a) and 3.10(b). Since maximum wavefront slope at 15-mm aperture is within the design baseline of the implemented 1-cm filter, the ODWS and UltraSurf measurements agree well, within 0.1λ RMS [Fig. 3.10(c)]. The Zernike coefficients (in fringe Zernike order) for these two measurements are in excellent agreement [Fig. 3.11]. The Zernike coefficients are obtained by fitting 37 Zernike polynomials, with terms 4 to 15 shown in Fig. 3.11. We note that the bump at the center of the wavefront is poorly fitted by Zernike polynomials, potentially leading to discrepancies.

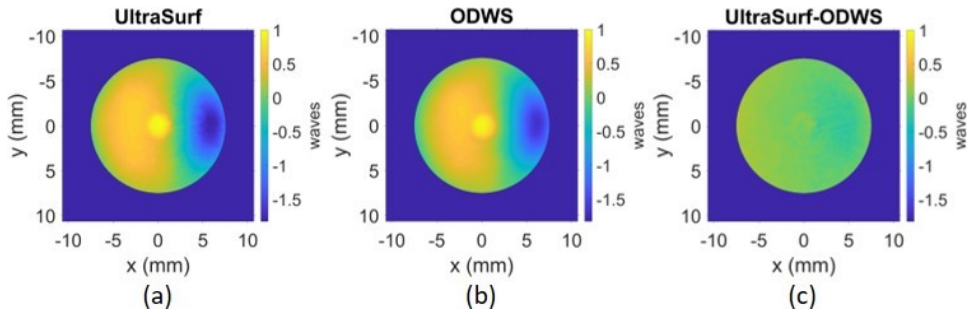

Figure 3.10 Wavefront of freeform phase plate #3 within 15-mm aperture (a) measured with UltraSurf, RMS = 0.66λ, PV = 2.8λ; (b) measured with ODWS, RMS = 0.59λ, PV = 2.61λ; (c) difference, RMS = 0.1λ.

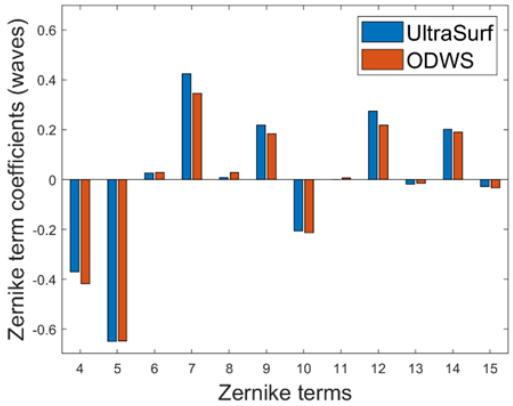

Figure 3.11 Comparison of Zernike coefficients of wavefront of phase plate #3.

When increasing the pupil size from 15 mm to 20 mm, the high wavefront slopes at the left edge of this phase plate exceed theoretical limit of the implemented ODWS ([-5.5, 19.3] vs. [-8, 8] waves/mm). The RMS difference of the measurements by UltraSurf and ODWS subsequently increases from 0.1λ to 0.16λ [Fig. 3.10(c) and Fig. 3.12(c)].

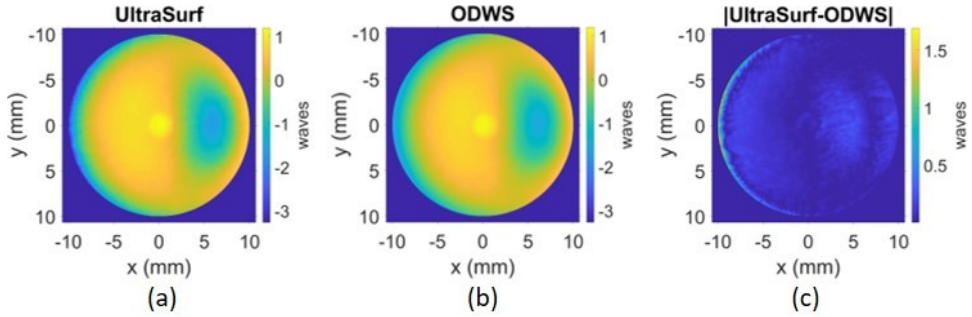

Figure 3.12 Wavefront of freeform phase plate #3 within 20 mm (a) measured with UltraSurf, RMS = 0.69λ, PV = 4.44λ; (b) measured with ODWS, RMS = 0.6λ, PV = 2.8λ; (c) absolute difference, RMS = 0.16λ.

In Fig. 3.12(c), the highest difference between the two measurements is observed at the left edge in the difference map (plotted as magnitude of the difference to emphasize the peak value on the left side of the aperture). This can be explained by fluence measurements performed in the far-field plane, where the ODWS filter is located, and in the detection plane, where the near-field fluence is measured. Figure 3.13(a) shows the far-field of phase plate #3 compared to the 1-cm filter width. The high-slope wavefront components that exceed the design baseline are clipped by the filter at the right edge, marked with a white line. This far-field clipping results in a dark patch

at the left edge in the fluence map of phase plate #3 at the detection plane as shown in Fig. 3.13(b). The resulting dark patch in fluence distribution due to clipping from the current filter-width design leads to less accurate wavefront reconstruction in that region, although good wavefront reconstruction is observed everywhere else.

We further demonstrate that this far-field clipping is the main source of inaccurate wavefront retrieval. Figures 3.13(c) and 3.13(d) show the measurement results through the same filter for phase plate #2 that has wavefront slopes close to the design baseline. The lower wavefront slopes lead to smaller spatial spread of the far field, well within the filter width. The wavefronts measured by the ODWS and the UltraSurf agree within 0.1λ RMS [Fig. 3.7(c)]

These results experimentally confirm that the ODWS dynamic range is effectively limited by filter width, because far-field components induced beyond that by large wavefront slopes lead to inaccurate wavefront reconstruction.

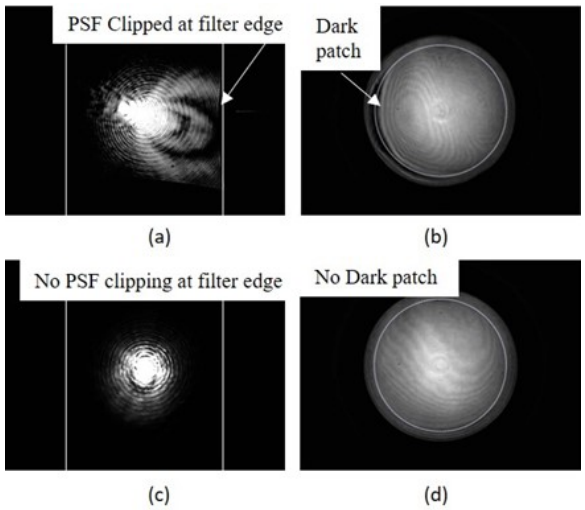

Figure 3.13 Fluence at far-field plane for (a) phase plate #3 and (c) phase plate #2, and at the detection plane for (b) phase plate #3 and (d) phase plate #2. All data is measured with a filter having 100% transmission over 1 cm. On (a) and (c), the ODWS-filter 1-cm width is shown by two white lines. On (b) and (d), the 20 mm measurement diameter is indicated by a white circle. The PSF images are intentionally saturated towards the center to make the edge visible.

Characterization of phase plate #3 with the Zygo interferometer was not successful as some portions in the center and on the edge of the wavefront could not be retrieved due to high slopes

there. A raw wavefront returned by the software is displayed in Fig. 3.14, showing areas where the wavefront was not reconstructed and set to NaN (Not a Number), same as background. The measurement has a resolution of 51.4 μm for a 22 mm measurement aperture.

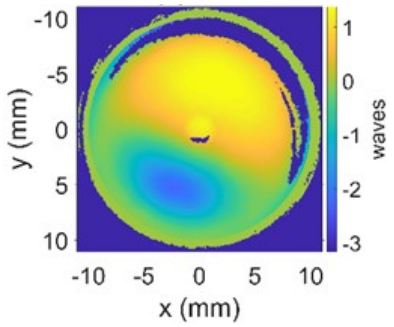

Figure 3.14 Wavefront of phase plate #3 measured with the Zygo interferometer.

3.5 Simulation study for impact of experimental imperfection and choice of optimal filter

The effect of experimental imperfections was simulated by considering an ODWS based on a 4*f* line with $f = 1$ m and detection with a camera having 7.7-μm pixels. This pixel size, which is set by simulation sampling requirements in the near field and far field, is close to the experimental pixel size (4.4 μm) and typical of scientific cameras. A wave with uniform fluence and phase given by a test wavefront is propagated to the far field, modulated by the filter transmission, and Fourier transformed back to the near field for detection. The fluences in the x and y direction are simulated, then used for wavefront reconstruction using the Southwell procedure. The test wavefronts correspond to Alvarez lens profiles proportional to $x^3/3+xy^2$ [48], with the quoted amplitude being the maximal amplitude over the 1-cm-diameter pupil. For a 1-wave amplitude, the largest wavefront slope is approximately 0.78 waves/mm in both the *x* and *y* direction over the 1-cm pupil. This indicates that the accuracy of an ideal ODWS is expected be excellent for these test wavefronts until their amplitude reaches approximately 10 waves, leading to wavefront slopes equal to $DR/2$ (8 waves/mm), for which some light in the far field reaches the edge of the filter.

Additive uncorrelated noise with standard deviation s normalized to the value of F_x and F_y for a flat input wavefront was introduced on all pixels of the two detected fluences F_x and F_y. Figure 3.15 shows the RMS error on the retrieved wavefront as a function of the test-wavefront amplitude for various values of s. It is clear that the ODWS can tolerate high levels of noise. The main effect of the detection noise is to decrease the dynamic range. In the absence of noise ($s = 0$), the RMS error increases sharply for large wavefront amplitudes which lead to energy at the edge of the far-field filter. When detection noise is present, the RMS error increases for low-amplitude test wavefronts. This is however a minor effect, considering that the largest noise level considered in these simulations (s=10%) is much larger than the typical noise of modern scientific cameras. For comparison, the RMS error obtained with three ideal pixelated binary filters (pixel size equal to 10 µm, 5 µm, and 2.5 µm) and noise-free detection has been plotted on the same figure. It is clear that noise due to filters with large pixels is the dominant factor limiting the accuracy for wavefront amplitudes approximately smaller than 5 waves, while photodetection noise becomes dominant beyond that. The induced errors remain below $\lambda/20$ over the ODWS dynamic range.

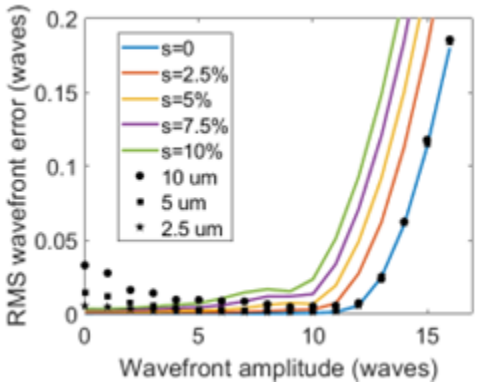

Figure 3.15 Calculated RMS error for the reconstructed wavefront of an Alvarez lens of varying amplitude for various noise levels from 0% to 10% and an ideal continuous filter (continuous lines). The RMS errors corresponding to noise-free detection with ideal pixelated filters (10 µm, 5 µm, and 2.5 µm) are indicated with markers.

We simulate the performance impact of the non-ideal filter transmission profile arising from fabrication uncertainties by two different approaches. The transmission profile (Eq. 2.3) was first modified using the following analytical formula:

$$t(u,v)=\frac{1}{2}+\frac{u}{W}+q\left[1-\left(\frac{2u}{W}\right)^2\right], \tag{3.1}$$

where the term proportional to q introduces a nonlinear transmission dependence relative to the far-field variable. Identically to the ideal transmission given by Eq. 2.3, the transmission given by Eq. 3.1 takes the values 0 and 1 at $-W/2$ and $W/2$, respectively, and has the slope $1/W$ at $u = 0$. The wavefront was retrieved using two wavefront slopes obtained from fluence distributions F_x and F_y simulated using the filter transmission profile given Eq. 3.1. Figure 3.16 shows the RMS error for various test-wavefront amplitudes and values of the nonlinear coefficient q.

We then fitted the measured transmission profiles of the three fabricated filters (pixel size equal to 10 μm, 5 μm, and 2.5 μm) with fourth-order polynomials to remove the noise associated with the transmission measurement and facilitate interpolation in the simulations. The determined polynomials were used to generate filter transmission profiles in the wavefront reconstruction simulations. Owing to the fact that filters are non-ideal, the field transmission is not linear over the filter width W, and the transmission slope is not uniformly equal to $1/W$ over the ODWS dynamic range. To compensate for this effect in the phase retrieval from the calculated F_x and F_y, the quantity W in Eqs. (2.11) and (2.12) was replaced by the inverse of the measured filter slope at $u = 0$. Using this effective slope compensates for the non-ideal filter transmission for wavefronts leading to far fields that are contained within the region where the filter transmission is linear with the identified slope. The simulated RMS reconstruction (Fig. 3.16) shows that the larger transmission profile error associated with smaller pixels leads to larger phase-retrieval error. While smaller pixels lead in theory to better accuracy because of lower noise due to pixelation,

transmission profile errors due to fabrication tolerance increase for smaller pixels. Comparison of Figs. 3.15 and 3.16 shows that the 2.5-μm filter is the optimal choice for wavefront smaller than a few waves, but the profile error leads to relatively larger measurement error for that pixel size than for larger pixel sizes for larger wavefronts. Reconstruction error smaller than $\lambda/10$ is achievable with the current filters, which were fabricated with a mild tolerance of +/-0.5 μm, over most of the ODWS dynamic range. Improving the accuracy and taking advantage of the smaller pixel size over the full dynamic range of the ODWS requires tighter fabrication tolerance to limit the transmission profile error to within a few percent. Digital halftoning algorithms generating pixel distributions that are less sensitive to fabrication errors could also be investigated [49].

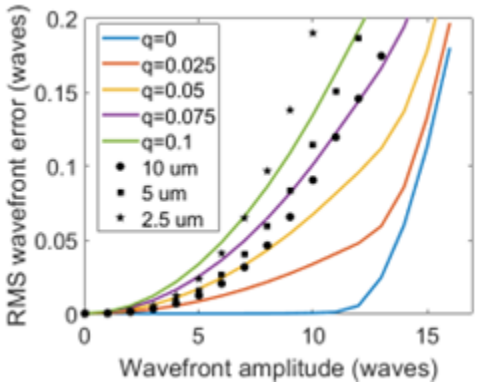

Figure 3.16 Calculated RMS error for the reconstructed wavefront of an Alvarez lens of varying amplitude for various nonlinearity coefficient q and an ideal continuous filter (continuous lines) and for the measured transmission profiles of the pixelated filters (markers).

Chapter 4 Miniaturization and system tolerance

In Chapter 3, we demonstrated ODWS based on $4f$ configuration. We used a down-collimating relay since the accuracy of an ODWS implemented with a spatially dithered distribution of transparent and opaque pixels depends on the ratio of the far-field size and pixel size. Smaller pixels generally provide better accuracy because they provide a transmission that is closer to the ideal linear profile. However, the pixel size is in general limited by fabrication technology and physical effects that can occur in components with features size of the order of the optical wavelength. The relative error on the pixel size, and therefore on the synthesized transmission profile, increases for smaller pixels because of fabrication tolerance. Increasing the focal length of the optical system in an ODWS (resulting in increased far-field spread) allows for an increase in precision with spatially dithered pixels, but such an increase can be impractical when using a standard $4f$ arrangement. A telephoto system can solve this issue by offering a large Effective Focal Length (EFL) in a relatively small footprint. While the larger EFL yields a smaller dynamic range, this can be compensated in practice by using larger width filters. In this work, we describe the design of telephoto based-ODWS that accurately characterizes freeform optics and significantly reduces the system length, and we evaluate its robustness to the misalignment of system components.

4.1 Design of Telephoto based ODWS

In a telephoto, the physical length from vertex to focal plane is less than the EFL. Figure 4.1 shows a telephoto system has an EFL of f (the distance between principal plane P_2 and focal plane F_2),

but the total length from vertex of first lens to focal plane is kf, where $k < 1$ is the telephoto ratio, defined as the ratio of distance between front vertex of the first lens and the focal plane of the telephoto with respect to the effective focal length of the telephoto. For a given separation $'d'$ between positive and negative lenses, EFL $'f'$ and telephoto ratio $'k'$, the focal lengths of two lenses required are: $f_1 = \frac{fd}{f(1-k)+d}$, $f_2 = \frac{(f_1-d)(kf-d)}{f_1-kf}$.

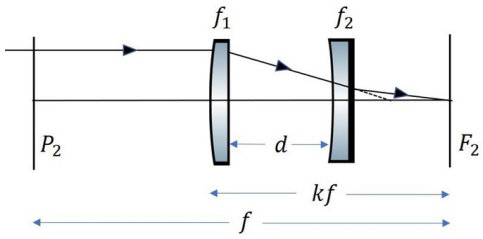

Figure 4.1 Layout of Telephoto lens.

We investigated the ODWS wavefront retrieval scheme i.e., the relationship between final intensity and wavefront slope. This also helps in understanding how the parameters will affect the wavefront measurement quality and the miniaturization. Figure (4.2) shows an ODWS system, where the first two lenses form telephoto. A gradient transmission filter is placed at the focal plane of a telephoto system with EFL of f. The wavefront to be tested is at the input plane at a distance d_0 from the vertex of first lens. The third lens having a focal length of f_3 is positioned at a distance a from the filter and it images the input plane to a detector at distance b.

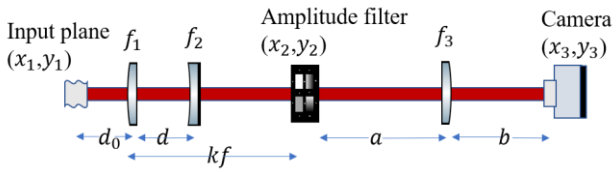

Figure 4.2 Telephoto based ODWS system layout.

Let $(x_1, y_1), (x_2, y_2)$ and (x_3, y_3) $r$epresent the coordinate planes of the test object location, filter location and detection plane respectively. Let $u_1(x_1,\ y_1)$ represent the field of the test object. The field at the filter location can be written by using the matrix method [50] as

$$u_2(x_2, y_2) = \frac{1}{j\lambda B} \iint_{-\infty}^{\infty} u_1(x_1, y_1) e^{\{j\frac{\pi}{\lambda B}[A(x_1^2+y_1^2)-2(x_1x_2+y_1y_2)+D(x_2^2+y_2^2)]\}} dx_1 dy_1, \quad (4.1)$$

where A, B and D are the elements of the $ABCD$ matrix of the propagation from object plane to filter plane, given by:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} 0 & f \\ -\frac{1}{f} & 1 - \frac{d}{f_2} - \frac{d_0}{f} \end{bmatrix}, \quad (4.2)$$

where f is given by $\frac{1}{f} = \frac{1}{f_1} + \frac{1}{f_2} - \frac{d}{f_1 f_2}$. Note that matrix element $B = f$, the EFL of the telephoto. Using $A = 0$ in Eq. (4.1), we get

$$u_2(x_2, y_2) = \frac{1}{j\lambda B} e^{j\frac{\pi D}{\lambda B}(x_2^2+y_2^2)} U_1(\xi, \eta), \quad (4.3)$$

where (ξ, η) are the spatial frequencies defined by $\xi = \frac{x_2}{\lambda B}\ and\ \eta = \frac{y_2}{\lambda B}$. $U_1(\xi, \eta)$ is the Fourier transform of input field. u_2 is not an exact Fourier transform of u_1, unlike in $4f$ based ODWS setup where an exact Fourier transform of the input field is obtained at the filter plane location.

The field at (x_2, y_2) is modified by a filter of linear transmission, $t_A = \frac{x_2}{W} + \frac{1}{2}$ where, W is the width of the filter. Using Eq. (4.3), the field at detection plane can be written as:

$$u_3(x_3, y_3) = \frac{-1}{(\lambda^2 BB')} \iint_{-\infty}^{\infty} \left(\frac{x_2}{W} + \frac{1}{2}\right) e^{j\frac{\pi D}{\lambda B}(x_2^2+y_2^2)} U_1(\xi, \eta) e^{\{j\frac{\pi}{\lambda B'}[A'(x_2^2+y_2^2)-2(x_2x_3+y_2y_3)+D'(x_3^2+y_3^2)]\}} dx_2 dy_2, \quad (4.4)$$

where A', B' and D' are the elements of the $A'B'C'D'$ matrix of the propagation from filter plane to detection plane through third lens, given by:

$$\begin{bmatrix} A' & B' \\ C' & D' \end{bmatrix} = \begin{bmatrix} 1-\frac{b}{f_3} & a+b-\frac{ab}{f_3} \\ -\frac{1}{f_3} & 1-\frac{a}{f_3} \end{bmatrix}. \tag{4.5}$$

Equation (4.4) can be simplified if the quadratic phase factors in coordinates (x_2, y_2) are cancelled, i.e., if the following condition is satisfied:

$$\frac{D}{B} = -\frac{A'}{B'}. \tag{4.6}$$

Using the matrix elements D, B, A' and B' from Eqs. (4.2) and (4.5) in Eq. (4.6) and solving for b, we get,

$$b = f_3 \frac{f_1 f_2 d_0 + f_1 f_2 d - f_2 d d_0 + (f_1 f_2 - f_1 d - f_1 d_0 - f_2 d_0 + d d_0)(kf - d + a)}{f_1 f_2 d_0 + f_1 f_2 d - f_2 d d_0 + (f_1 f_2 - f_1 d - f_1 d_0 - f_2 d_0 + d d_0)(kf - d + a - f_3)} \tag{4.7}$$

If we calculate the conjugate location of the phase object in ODWS by calculating its image position through each lens, we arrive at the expression in Eq. (4.7). Therefore, if the detector is at imaging position of phase object, then Eq. (4.6) is satisfied. To solve Eq. (4.4), we invoke the Fourier derivative property which can be written as

$$\frac{\partial g(x_1, y_1)}{\partial (x_1)} = \iint_{-\infty}^{\infty} j2\pi\xi G(\xi, \eta) e^{j2\pi(x_1\xi + y_1\eta)}\, d\xi d\eta, \tag{4.8}$$

where G and g are Fourier transform pairs. With the detector being at image plane of test object and using the Fourier derivative property, we can write Eq. (4.4) as:

$$u_3(x_3, y_3) = e^{j\frac{\pi D'}{\lambda B'}(x_3^2 + y_3^2)} \left[\frac{j\lambda B^2}{2\pi W B'} \frac{\partial u_1\left(\frac{-x_3 B}{B'}, \frac{-y_3 B}{B'}\right)}{\partial\left(\frac{-x_3 B}{B'}\right)} - \frac{B}{2B'} u_1\left(\frac{-x_3 B}{B'}, \frac{-y_3 B}{B'}\right) \right] \tag{4.9}$$

Assuming the input field amplitude A_0 to have negligible variation and phase distribution to be φ that we want to characterize, i.e.

$$u_1(x_1, y_1) = A_0 e^{j\varphi(x_1, y_1)}, \tag{4.10}$$

Eq. (4.9) can be simplified to obtain:

$$u_3(x_3, y_3) = e^{j\frac{\pi D'}{\lambda B'}(x_3^2 + y_3^2)} e^{j\varphi\left(\frac{-x_3 B}{B'}, \frac{-y_3 B}{B'}\right)} \left(\frac{-A_0 B}{B'}\right) \left[\frac{\lambda B}{2\pi W} \frac{\partial \varphi\left(\frac{-x_3 B}{B'}, \frac{-y_3 B}{B'}\right)}{\partial\left(\frac{-x_3 B}{B'}\right)} + \frac{1}{2} \right]. \tag{4.11}$$

The corresponding fluence at detection plane is given as

$$F_x(x_3, y_3) = \left(\frac{-A_0 B}{B'}\right)^2 \left[\frac{\lambda B}{2\pi W} \frac{\partial\varphi\left(\frac{-x_3 B}{B'}, \frac{-y_3 B}{B'}\right)}{\partial\left(\frac{-x_3 B}{B'}\right)} + \frac{1}{2}\right]^2. \tag{4.12}$$

Field at (x_3, y_3) with filter of transmission $t_A = 1$ can be obtained similarly and the corresponding fluence at detection plane in this case is

$$F_0(x_3, y_3) = \left(\frac{-A_0 B}{B'}\right)^2. \tag{4.13}$$

From Eq. (4.12) and (4.13), we obtain,

$$\frac{\partial\varphi\left(-\frac{x_3}{m}, -\frac{y_3}{m}\right)}{\partial\left(-\frac{x_3}{m}\right)} = \frac{\pi W}{\lambda f}\left(2\sqrt{\frac{F_x(x_3, y_3)}{F_0(x_3, y_3)}} - 1\right), \tag{4.14}$$

Eq. (4.14) calculates wavefront slope at input plane location from the fluences at detection plane. The negative sign is due to inversion of the field. $f = B$ is the EFL of the telephoto lens pair. $m = \frac{B'}{B} = \frac{a+b-\frac{ab}{f_3}}{f}$ is the magnification between input and detection plane. When the system is aligned to be afocal, i.e., $a = f_3$, the magnification *m* is simply $\frac{f_3}{f}$. For *m*=1, Eq. (4.14) describes an ODWS based on two lenses of identical focal length *f* set in a 4*f* configuration. Similarly, the wavefront slope along the *y* direction can be obtained from the detected near-field fluence after rotating the far-field filter by 90 degrees. The dynamic range for wavefront slope measurement is set by the filter's width and EFL of the telephoto system, i.e., $DR = 2\pi W/\lambda f$ in units of radians per unit length ($W/\lambda f$ in units of number of waves per unit length).

4.2 Experimental demonstration using phase plate measurement

The experimental setup (Fig. 4.3) closely resembles the system layout presented in Fig. 4.2. Lenses 1, 2 and 3 are commercial off-the-shelf lenses (Thorlabs), chosen as plano-convex, plano-concave

and plano-convex, respectively. The collimated input laser beam (λ = 633 nm) with a diameter of 20 mm propagates through a freeform phase plate which introduces a wavefront distortion. The beam then propagates through the telephoto system and is focused onto the transmission filter. The modulated beam propagates through the third lens and is captured on the camera (@The Imaging Source, DMK 23U274) located at the image plane of the phase plate. The system design parameters are shown in Table 4.1 referring to the layout in Fig. 4.2.

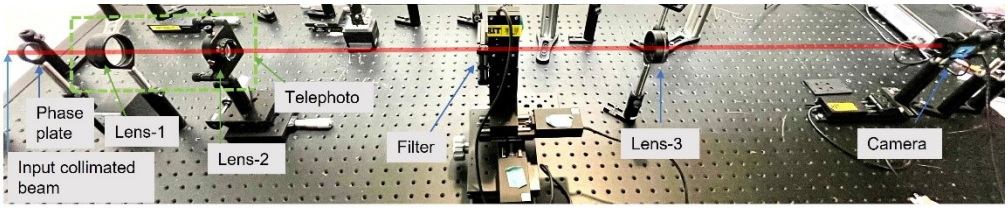

Figure 4.3 Telephoto ODWS experimental system.

Table 4.1 Design parameters for an ODWS implemented with a telephoto system.

Object distance from Lens-1, d_0	140 mm
Lens-1 focal length, f_1	200 mm
Lens-2 focal length, f_2	-25 mm
Distance between Lens-1 and Lens-2, d	176.43 mm
Distance between Lens-2 and filter plane, $kf - d$	412.1 mm
Lens-3 focal length, f_3	300 mm
Distance between filter plane and Lens-3, a	300 mm
Detecting plane distance from Lens-3, b	506.6 mm
Telephoto EFL, f	3.5 m

The experiments in this work were performed using the pixelated filter with pixel size of 2.5 μm and width W equal to 1 cm. The range of slopes that can be measured around the mean slope is [-2.3, 2.3] waves/mm, and the dynamic range is 4.6 waves/mm. Accurate wavefront reconstruction requires optimum far-field spot size that is proportional to the EFL so that the sampling from the binary pixelated filter is sufficient. A telephoto EFL ~3.5 m was chosen based on different experimental trials to reconstruct a plane wavefront accurately, i.e., when no aberration is induced on the input wave with a phase plate.

Referring to the layout in Fig. 4.2, for convenience in determining the experimental telephoto EFL, the system was aligned to be afocal, i.e., the distance between the filter and Lens

3 is $a = f_3$. In this case, the magnification is $m = \frac{f_3}{EFL}$, and the experimental EFL can be determined from the measured magnification without the need to determine *a* or *b* accurately. Both f_3 and EFL determine the magnification of the system and hence the measurement resolution. The parameters of the optical system determined experimentally are m = 1/11.4044, EFL(f) = 3421.3 mm, yielding a spatial resolution in the input plane equal to 50.2 μm based on the 4.4-μm camera pixel size.

The total length of this telephoto-ODWS system, from object to detection plane is ~1.5 m. For a similar EFL and measurement resolution using the same camera, the length of an ODWS without a telephoto ($f_1 = 3.5$ m, $f_3 = 300$ m) would be 7.6 m. In this demonstration, the use of a telephoto system reduces the footprint by 5 compared to a 4*f*- ODWS system. More compact implementations with longer EFL are possible following the same concept.

In order to assess the performance of the telephoto-lens-based ODWS, we have measured the wavefront induced by two of the phase plates as shown in Chapter 3. Each phase plate was previously measured by Ultrasurf with a spatial resolution of 250 μm. The UltraSurf data are interpolated to match the ODWS data resolution. A reference wavefront without any phase plate is measured with the ODWS to account for static aberrations present in the system. This reference wavefront is subtracted from each test wavefront. The difference in the wavefronts reconstructed by the ODWS and the UltraSurf is taken as a metric for accuracy validation, after numerical removal of the piston and tip/tilt.

The measurements of phase plate #1 by UltraSurf, ODWS and their comparison are shown in Figs. 4.4(a), (b) and (c), respectively. Similarly, the measurements of phase plate #2 are shown in Fig. 4.5. Each ODWS measurement is an average of 10 measurements. For lateral registration of the wavefronts, the centroid of the peak at center of the phase plates is used as fiducial. For

clocking alignment, each wavefront is rotated in a range of angles and for each angle, Zernike polynomials are fit to the wavefront. The wavefronts primarily have coma aberrations, and we choose coma in the horizontal direction as reference. The angle that gives the minimum Zernike coefficient for horizontal coma is chosen for numerical rotation before calculation of the wavefront difference. The wavefront alignments were done separately for UltraSurf and ODWS measurements in MATLAB. In future, fiducials can be machined on the parts for wavefront alignment and easier comparison.

The telephoto-lens-based ODWS wavefront measurements are comparable with UltraSurf measurements; their RMS difference is 0.09λ for both phase plates. This difference may be caused by imperfect registration and may include inherent error in either measurement technique. The ODWS measurements for each phase plate show an average precision of ~λ/50, quantified as the RMS of difference between each wavefront map and the average of all the measurements.

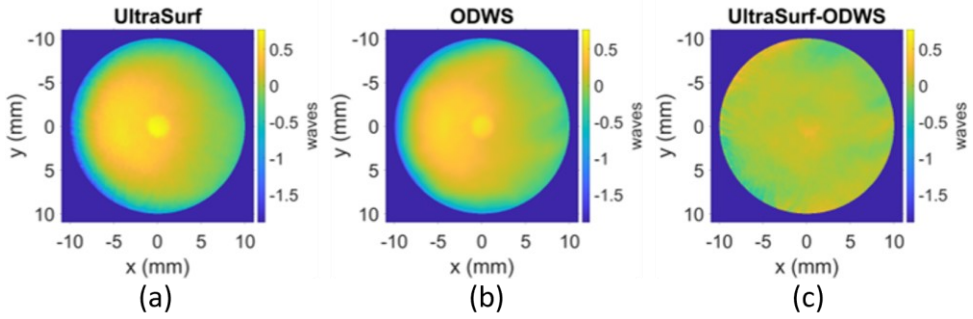

Figure 4.4 Wavefront of phase plate #1: (a) UltraSurf measurement [PV: 2.66λ, RMS: 0.4λ]; (b) ODWS measurement [PV: 2.32λ, RMS: 0.37λ]. (c) Difference (RMS 0.09λ).

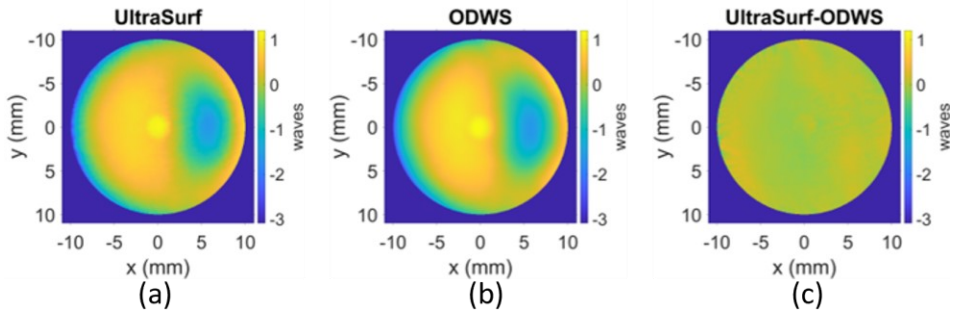

Figure 4.5 Wavefront of phase plate #2: (a) UltraSurf measurement [PV: 4.25λ, RMS: 0.67λ,]; (b) ODWS measurement [PV: 3.8λ, RMS: 0.7λ]; (c) Difference (RMS: 0.09λ).

4.3 Alignment tolerance

In this section, we analyze the error in wavefront measurement due to misalignment of the ODWS components along the optical axis. First, we give a description of the alignment of the telephoto ODWS system. Referring to Fig. 4.2, the three lenses are initially put at their respective positions according to the design in Table 4.1 and are optically aligned for tilting and centering. An object with known dimension is placed in front of Lens 1 at the design object distance and is imaged to a camera, from which the object location and magnification of the system are determined. The amplitude filter, mounted on a three-axis translation stage, is placed at the focal plane of the telephoto system. Its longitudinal position is determined by a knife edge test, using the sharp transition from opaque to transparent transmission at the edge of the filter. The filter is centered with respect to the beam using fiducials added at the four edges of the filter. During the alignment, a Shack-Hartman wavefront sensor is used after Lens 3 to ensure that the system is afocal. The exact EFL of telephoto is determined by dividing the focal length of Lens 3 by the magnification.

Referring to Fig. 4.2, the alignment uncertainties are primarily from the position of the object, filter, and detector. Hence for tolerance analysis, each of these components was separately moved from its ideal position over a range of [-8 mm, +8 mm], in steps of 2 mm (only one

component was moved at a time while the other components were kept at their ideal positions). These longitudinal translations are facilitated by motorized translation stages.

Figures 4.6(a) and (b) show the wavefront measurement error in relation to variation of the object and filter position respectively. The measurement error is quantified as the RMS of the difference between wavefront measured at misaligned position and wavefront at ideal position. These figures show that the ODWS is tolerant to object and filter position misalignments. The maximum RMS error with respect to ideal position is within $\lambda/10$ for both phase plates. Considering that the range of misalignments ([-8 mm, +8 mm]) is significantly larger than the typical uncertainty during alignment, we conclude that this system can tolerate typical misalignments of the object and filter positions. Figure 4.7 shows that the wavefront RMS error also increases with the misalignment of the detector. It is within $\lambda/12$ and $\lambda/20$ for a misalignment range of [-8 mm, 8 mm] and [-2 mm, 2 mm], respectively.

Figures 4.6 and 4.7 also show that the measurement of wavefronts with steeper slopes is more sensitive to misalignments of the object, filter and detector. Phase plate #2 has higher wavefront amplitude (1.6 times) and steeper slope (3.2 times for the maximum slope) at edge compared to phase plate #1. It therefore has higher error than phase plate #1. It is noted that the detector misalignment data were obtained with a reduced pupil diameter of 17 mm to compare the results for the same misalignment range. For a full 20-mm beam, the wavefront measurements were not valid when misalignment is larger than +4 mm as the much higher wavefront slope at the edge diffracts the beam away.

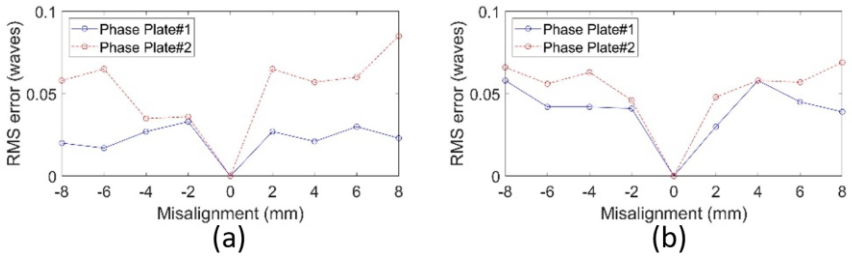

Figure 4.6 RMS difference of wavefront measurements from ideal position when varying (a) the object position, and (b) the filter position.

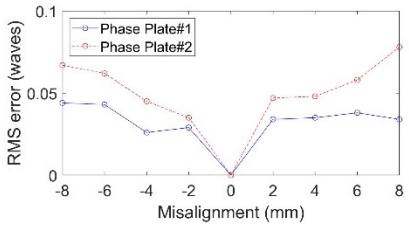

Figure 4.7 RMS difference of wavefront measurements from ideal position when varying the detector position.

The presented tolerance analysis was performed for component misalignments along the optical axis. Object and detector position misalignments in the lateral directions result in transverse shifts of the measured fluence profiles. This can be handled during the registration process after wavefront reconstruction from the measured fluences. Transverse translation of the filter in the direction orthogonal to its gradient has no impact on the ODWS operation because this does not modify the modulation induced on the field. Translation of the filter in the direction of the gradient introduces a global tilt on the reconstructed wavefront. Such tilt is not a concern for components metrology.

We have also analyzed the wavefront sensing performance of the telephoto ODWS system @Zemax optical design software to estimate the effects of tip/tilt misalignments. The peak-to-valley (PV) and RMS errors in retrieving a test wavefront in the nominal system are 0.0886λ and 0.0146λ respectively. For 0.8-degree tip and tilt misalignment in filter, the respective PV and RMS errors are 0.0889 waves and 0.0148 waves. For 0.8-degree tip and tilt misalignment in camera, the PV and RMS errors are 0.1052 waves and 0.017 waves. The experimental tilt/tip

alignment uncertainty in the filter and camera plane is less than 0.1°, much less than the modeled tilt/tip effected. Therefore, the ODWS measurements are tolerant to tip/tilt alignment error of these two components.

Chapter 5 Wavefront sensing using deep learning in nonlinear filter based ODWS

5.1 Need for nonlinear filter

We discussed the trade-off between dynamic range and sensitivity in Section 2.5. A step like filter will have maximum sensitivity, this is like a pyramid wavefront sensor. A filter of large width can give high dynamic range. A qualitative illustration done in [42] is shown here. Figure 5.1 shows the amplitude transmission of filters of three different widths. The spherical phase object in the simulation has a far-field of width corresponding to the width of filter-2.

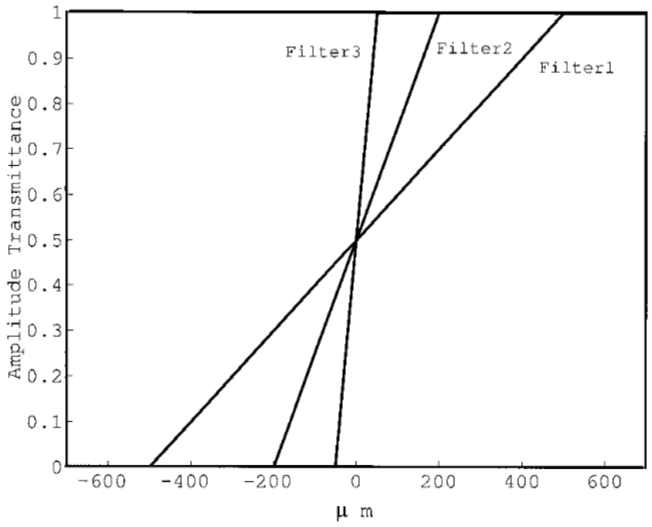

Figure 5.1 Amplitude transmittance of three linear filters. Figure credit [42].

Figure 5.2 shows the corresponding fluence images. The contrast is weak for Filter 1 which is the low sensitivity case. The contrast is good in the case of Filter 2 because its width corresponds matches the far-field. However, Filter 3 is narrower than the far-field; so, the contrast is partially enhanced but the image is saturated for the far-field region outside Filter 3 decreasing the dynamic range.

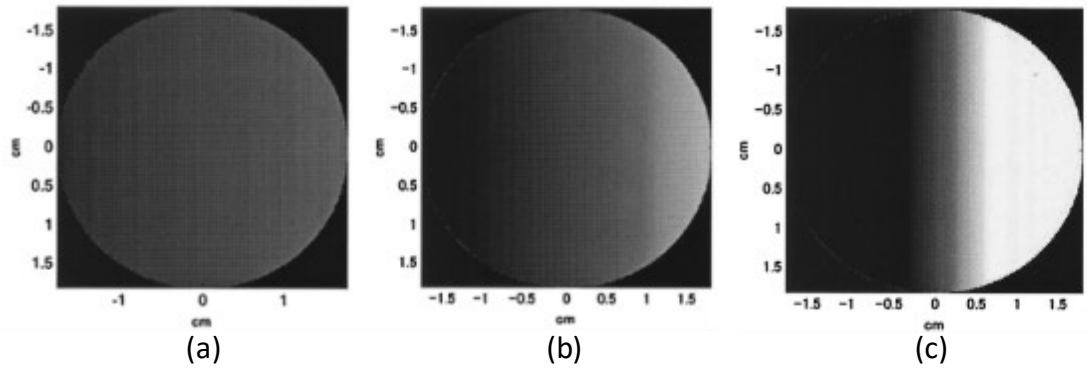

Figure 5.2 Calculated fluence images for (a) filter 1, (b) filter 2, (c) filter 3 in Fig. 5.1. Figure credit [42].

By combining both the step and linear filter, the trade-off between dynamic range and sensitivity can be controlled by balancing the relative power of two extreme filters. A nonlinear transmission profile with relatively high amplitude transmission gradient near the center of a large filter width thus is both sensitive and high dynamic range. The transmission profiles of the binary pixelated nonlinear filters investigated in this work are described by [39]

$$T_x(x,y) = \left(\frac{1-\beta}{2}\right)\left(1+\frac{x}{W/2}\right) + \frac{\beta}{1+e^{-\frac{2x}{\sigma W}}}. \tag{5.1}$$

It is a combination of a linear and a step function where β and σ determine the height and width of the step respectively. In this work, we investigate the CNN with three filters (one linear and two nonlinear). Figure 5.3 shows the designed transmission and corresponding slope profiles where the transmission gradient is in x-direction. The three filters of specifications in Table 5.1 were designed using the binary pixelated method with 2.5μm size pixels. A design example is shown in Fig. 5.4. The nonlinear filters use steep and shallow transmission slopes for low and high spatial frequencies respectively. Such general filter profiles can be realized by spatial dithering and binary pixelation, however, it has not been investigated for ODWS. When such a non-linear transmission modulates the far field, a closed-form analytical solution to relate the fluence images

to wavefront slopes cannot be obtained in general, due to infinite number of terms arising from the series expansion of a non-linear profile.

Table 5.1 Amplitude transmission parameters of one linear and two nonlinear filters.

σ	β	Transmission Slope at center in units of $(mm)^{-1}$
	0	Linear filter, slope = 0.05
0.1	1/4	0.1
	3/4	0.2

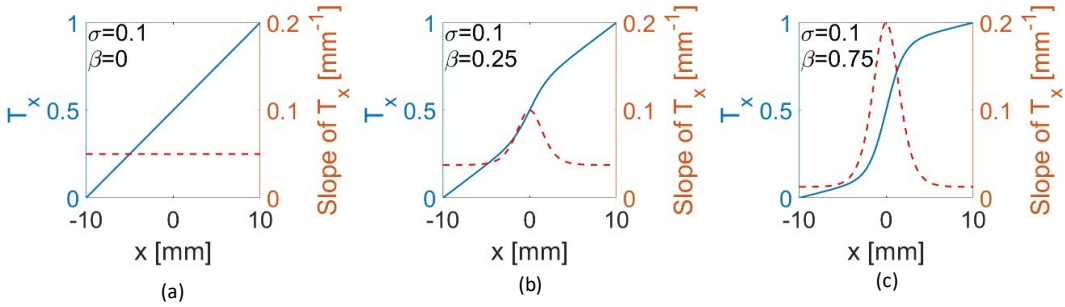

Figure 5.3 Amplitude transmission and its slope profiles of (a) linear, LF, (b) nonlinear filter 1, NLF1, and (c) nonlinear filter 2, NLF2.

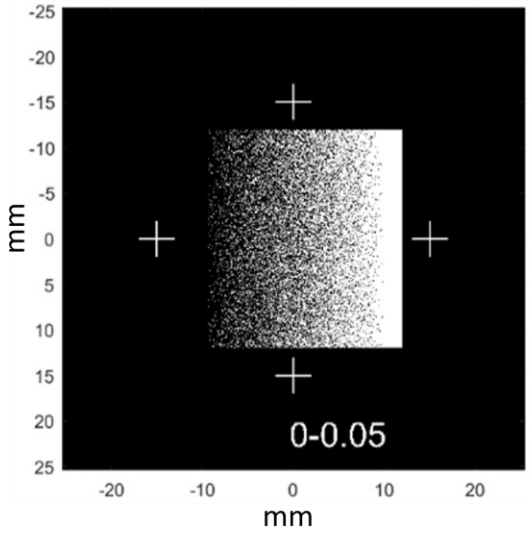

Figure 5.4 A design example of binary pixelated non-linear ODWS filter.

5.2 Investigation of phase retrieval

We explored iterative phase retrieval method to retrieve wavefront from fluence measurements in ODWS with non-linear filter. It was inspired by the Gerchberg-Saxton algorithm [51,52]. In phase retrieval, the goal is to recover the unknown input phase from the ODWS fluence measurements for two orthogonal filter transmissions. The input to the algorithm is the amplitudes of electric fields at the object plane, filter plane and the image plane. The amplitudes are proportional to the square roots of the measured fluences. The algorithm can be described in the following steps. An initial estimate of the phase at object plane is done by generating a random phase array. This estimated phase along with the measured amplitude forms an estimate of the complex electric field at object plane. This field is numerically propagated through ODWS and fluence maps are obtained for horizontal gradient filter. The phase computed from the complex field at image plane is combined with the measured amplitude. This forms an estimate of the field at image plane for the horizontal filter gradient. This field is back propagated to get an estimate of the field at filter plane. To get the field just before the filter, the amplitude of the field at back of the filter was replaced to be the square root of generated fluence there. This field was again propagated to the object plane. The phase computed from this field at object plane is combined with the measured amplitude and the aperture constraint, to form a new estimate of the field at object plane. This estimated input field is propagated to image plane and then back to the object plane though vertical gradient filter in same way to get a new estimate of the input field. The extracted phase at object plane is the input to the algorithm for next iteration. After 100s of iterations, we compared the retrieved input phase with the ground truth phase we took as input for data generation.

The algorithm was tested for linear and non-linear filter based ODWS for different wavefronts. The results imply that the convergence of the algorithm depends on both the wavefront shape and initial guess. The challenge appears to be in the convergence to a solution when the initial guess is not close to the solution and in the case of complex wavefront shape. Another potential challenge is dividing the filter transmission to get the field just before the filter during backpropagation from camera to input plane. The algorithm worked in some cases where the aperture constraint is applied and without this constraint the algorithm for same situation did not converge to the solution. Here the input amplitude profile used was gaussian; the algorithm should also be investigated for top-hat input amplitude profile.

Next, we decided to investigate an alternative method of using machine learning for wavefront reconstruction in ODWS.

5.3 Convolutional Neural Network for ODWS wavefront reconstruction

Convolutional Neural Network (CNN)-based approach has attracted interest in image-based wavefront sensing [53-55]. This data-driven deep learning approach allows to establish the complicated nonlinear relations between system inputs and outputs, without knowing the details of a physical model of the system. Therefore, it is especially useful in image analysis tasks. CNN architecture can be designed to predict spatial phase information [55-58], or coefficients of basis functions such as Zernike polynomials and deformable mirror modes [41,53]. The deep-learning-based wavefront estimation method has been demonstrated to be robust in the presence of noise and misalignments in deflectometry and Shack Hartman sensing methods [56,58].

For robust phase reconstruction of generally shaped wavefronts (that cannot be described by modal coefficients), zonal reconstruction is preferred. This is because it is difficult to accurately

train and predict such wavefronts with a CNN architecture that has modal coefficients as output [55,58].

This chapter presents our investigation on CNN for robust phase reconstruction in ODWS based on a binary pixelated filter. Figure 5.5 shows the CNN architecture implemented to retrieve wavefronts from the ODWS fluence images. The input and output of the CNN are the fluence image ratios and corresponding reconstructed input wavefront respectively. The CNN is inspired by the U-Net architecture [55-59]. It is trained using mini batches of data.

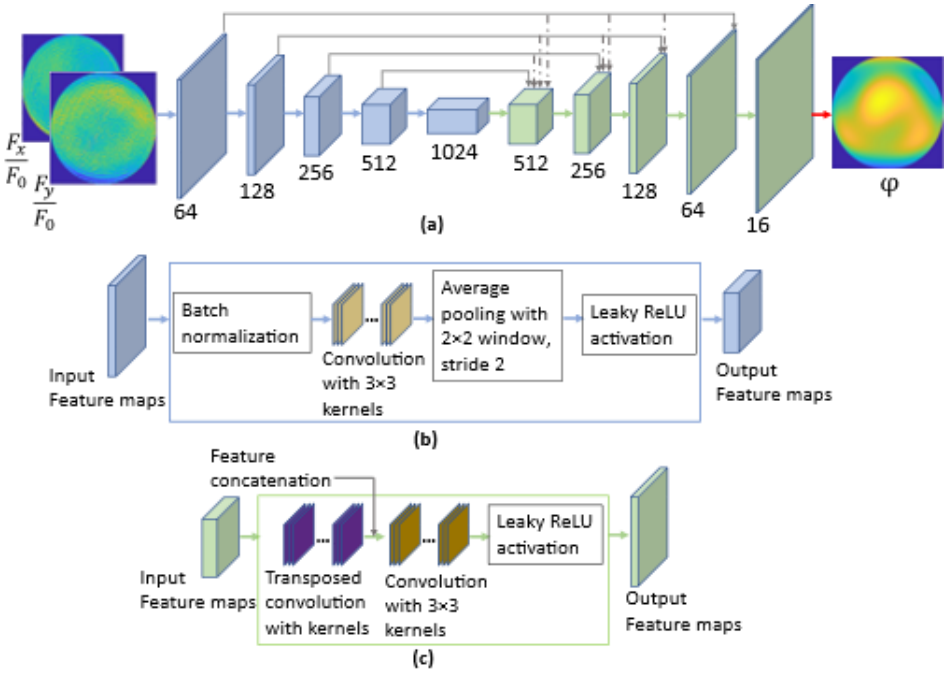

Figure 5.5 Overall CNN architecture with encoder (blue blocks) and decoder (green blocks) layers for wavefront reconstruction from ODWS fluence ratio maps; the number under each block is the depth of the feature maps or the number of kernels in that layer. (b) Architecture of an encoder layer. (c) Architecture of a decoder layer.

The two-channel input (fluence image ratios) of all training samples in the mini batch are processed by a series of encoder layers. In Fig. 5.5(a), we have five of the encoder layers, shown in blue blocks. Architecture of an encoder layer is shown in Fig. 5.5(b). Let us first consider the case of a single training sample, meaning a batch size of one. The convolutional layer in a neural

network performs a convolution operation on the input data using learnable filters/kernels. The mathematical representation of this operation is as follows:

$$O_{i,j,k} = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \sum_{l=0}^{L-1} I_{(i+m),(j+n),l} \cdot K_{m,n,l,k} + b_k \quad (5.2)$$

where:

- $O_{i,j,k}$ is the element at position (i, j, k) in the output feature map.
- M and N are the spatial dimensions of the kernel K. Specifically, we use $M = N = 3$.
- L is the number of channels in the input tensor I.
- $I_{(i+m),(j+n),l}$ is the element at the position (m, n, l, k) in the kernel K. Note that each kernel operates on all of the L channels of the input tensor.
- b_k is the bias term for the k-th kernel.
- The numbers under the blocks (64, 128…) represent the number of kernels (and hence the number of output features/channels, also called the **depth**) of that layer.

The output feature maps of the convolutional layer are then processed through the average pooling. For each 2x2 square in the input feature map, we calculate the average of the numbers in that square. However, instead of moving the 2x2 window one step at a time, we move it two steps at a time. This skipping is called the "stride". The result is a downsampled output feature map. This operation reduces the height and width of the feature maps by factor of two. Pooling layers are used to reduce the dimensions of the feature maps by summarizing the features. Hence, it decreases the number of parameters to be learned and the computational workload within the network.

The output from the pooling layer is then processed through activation function. The purpose is to introduce non-linearity so that CNN can learn complex mapping. Here we use leaky rectified linear unit (ReLU) activation. Mathematically, it can be written as follows:

$$f(x) = \begin{cases} \alpha x, & if\ x < 0 \\ x, if\ x \geq 0 \end{cases} \tag{5.3}$$

where $\alpha = 0.01$ is small positive constant. With each consecutive encoder layer, the spatial size (height and width) of the feature maps is halved by the average pooling, but the number of feature maps are doubled, because the number of kernels is being doubled (64, 128...).

After the encoder layers, the feature maps are processed through the decoder layers (shown in green blocks). Figure 5.5(c) shows the architecture of a decoder layer. The spatial dimensions of the input feature maps are expanded by applying transposed convolutions. Unlike standard convolution, which decreases input elements through the kernel, transposed convolution broadcasts input elements through the kernel, resulting in an output that is larger than the input. As an example, Fig. 5.6 shows how the transposed convolution of a 2x2 input tensor with a 2x2 kernel is computed to produce a 3x3 output. The weights in the kernel are learnable during the training process.

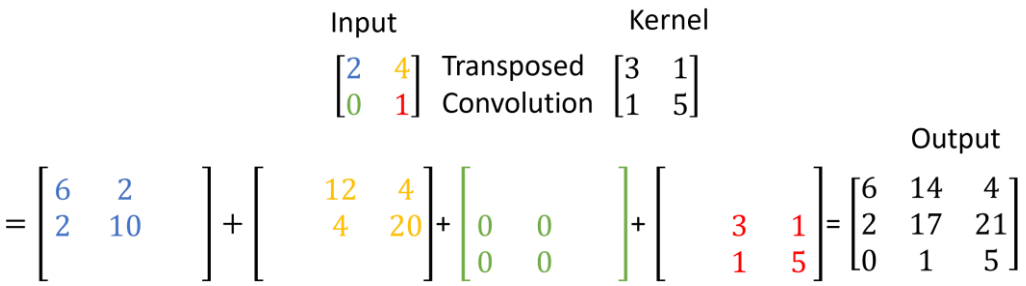

Figure 5.6 Example showing transposed convolution operation.

The spatial dimensions of the input feature maps are doubled by applying transposed convolutions in the decoder layers with appropriate kernel size and stride, while reducing the depth by half. For example, the first decoder layer takes in 1024 features and by setting the number of transpose convolution kernels to 512, the depth of its output is halved. Then the feature maps from the output of different encoder layers are concatenated to the decoder layers (gray arrows) to enhance the wavefront reconstruction details. For concatenation, features maps from the encoder

layers are first matched to the spatial dimension in decoder layers by average pooling with appropriate pooing window and stride. For example, the feature concatenation in the first decoder layer produces feature maps of depth 1472 (512 from the transposed convolution and concatenation from the encoder layers of depth 64, 128, 256, 512 respectively). The feature concatenation is followed by convolution (3×3 kernel) and leaky ReLU operations; here the number of convolution kernels is set to produce feature maps of depth equal to the depth of the output of the prior transposed convolution (except for the last decoder layer).

The last decoder layer in Fig. 5.5(a) takes in 64 features maps, the transposed convolution operation in it produces 32 feature maps. In this layer there is no concatenation from the encoder layers and the convolution operation is set to produce 16 output feature maps or depth. The output of this last decoder layer is resized to depth of one by applying 1×1 convolution (indicated as the operation in red arrow in Fig. 5.5a). The output represents the reconstructed wavefront map which spatial array dimension is same as that of the input fluence ratio maps.

These weights and biases of all the kernels in all layers of the CNN represent the mapping from input to output, and their optimal values would be realized through the training process. CNN is trained using mini batches of training samples to achieve faster learning. As shown in Fig. 5.5(b), the batch input to the encoder layer is first processed through batch normalization. It scales the feature maps of the mini-batch by the mean and standard deviation of the mini-batch. This step reduces the problem of the input values changing. It makes them more stable, making the model more general.

We implemented the CNN in the PyTorch framework [60]. The weights and biases are randomly initialized and then optimized during the training. CNN loss function in the form of RMS of residual between the ground truth and CNN predicted wavefront is computed. The loss in

essence is a function of weights and biases of the CNN. The loss is minimized using the Adam optimizer which is based on stochastic gradient descent algorithm. The optimizer updates the weights and biases of the CNN based on the gradients of the loss function. A learning rate of 0.0001 is used that determines the rate of training convergence.

5.4 Method for beam propagation simulation

Training CNN requires a large variety of wavefronts and corresponding ODWS fluence images. A numerical model was developed to predict fluence images and to simulate the theoretical wavefront measurement performance of our experimental ODWS system, shown in Fig. 5.7. The pixel size of the realized filter is 2.5-µm.

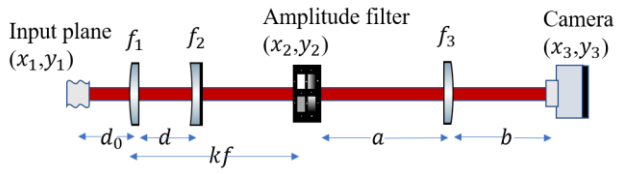

Figure 5.7 Telephoto based ODWS system layout.

Table 5.2 Design parameters for an ODWS implemented for CNN training.

Object distance from Lens-1, d_0	700 mm
Lens-1 focal length, f_1	200 mm
Lens-2 focal length, f_2	-25 mm
Distance between Lens-1 and Lens-2, d	176.43 mm
Distance between Lens-2 and filter plane, $kf - d$	412.06 mm
Lens-3 focal length, f_3	300 mm
Distance between filter plane and Lens-3, a	300 mm
Detecting plane distance from Lens-3, b	502.24 mm
Telephoto EFL, f	3.5 m

A beam propagation model was developed in MATLAB considering the ODWS design parameters as per Table 5.2. The model input considers a top-hat beam of diameter 9.73 mm. The input beam is aberrated by pre-specified wavefronts. The simulation grid size is (Mx $M = 2^{15}$x 2^{15}) with sampling interval of $\Delta x_1 = 1.25\mu m$, the array length, $L_1 = M\Delta x_1 = 41mm$,

approximately being four times the wavefront measurement aperture which covers ~7780×7780 samples. The sampling interval is chosen to have four samples (two in one direction) for each of the 2.5µm size pixel of the binary pixelated filter. The highest spatial frequency that can be modelled is the Nyquist frequency $f_N = 1 \,/\, (2\Delta x_1) = 400 \frac{cycles}{mm} or \left(400 \frac{waves}{mm}\right)$. The spatial frequency coordinate samples along one axis are $f_{x_1} = \frac{-1}{(2\Delta x_1)} : \frac{1}{L_1} : \frac{1}{(2\Delta x_1)} - \frac{1}{L_1}$, corresponding to the source plane coordinate samples, $x_1 = \frac{-L_1}{2} : \Delta x_1 : \frac{L_1}{2} - \Delta x_1$.

Free-space beam propagation from a source plane (field, U_1) to an observation plane (field, U_2) at a distance 'z', as depicted in Fig. 5.8, can be obtained by using the Rayleigh-Sommerfeld diffraction:

$$U_2(x,y) = \iint U_1(\xi,\eta) h(x-\xi, y-\eta) dx dy \tag{5.4}$$

where $h(x,y) = \frac{z}{j\lambda} \frac{e^{jkr_{12}}}{{r_{12}}^2}$, $r_{12} = \sqrt{z^2 + x^2 + y^2}$ and k is the wave number.

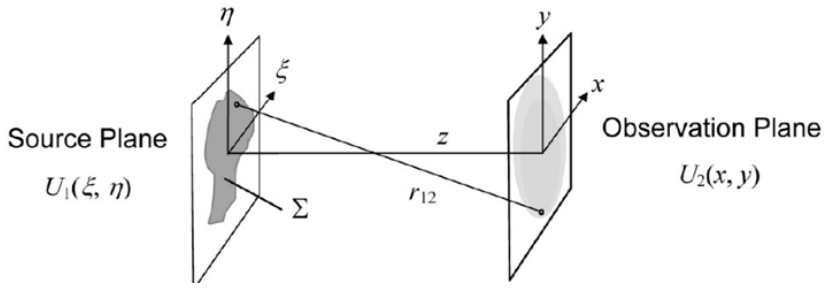

Figure 5.8 Propagation geometry for source and observation planes. Figure credit [61].

Using Fourier convolution theorem, Eq. (5.4) can be expressed as,

$$U_2(x,y) = \mathcal{F}^{-1}\left[\mathcal{F}\{U_1\} \times H\left(f_x, f_y\right)\right], \tag{5.5}$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ are the Fourier and inverse Fourier transform operation respectively. $\left(f_x, f_y\right)$ are the spatial frequency coordinates corresponding to the spatial coordinates (x, y). H is the transfer function given as

$$H(f_x, f_y) = e^{jkz\sqrt{1-(\lambda f_x)^2-(\lambda f_y)^2}} \tag{5.6}$$

The beam propagation model considers the effect of lenses by its phase transmission given as $e^{-j\frac{k}{2f}(x^2+y^2)}$, where f is the focal length of the lens.

We designed the binary pixelated filters with width of 20mm and a pixel size $2.5\mu m$ and other parameters as shown in Table 5.1, using the error diffusion algorithm (Section 2.4). Each pixel is then resampled to $1.25\mu m$ by assigning the filter design. This resampled transmission profile is loaded into the simulation, which modulates the far-field. One unmodulated and two modulated (for orthogonal filter orientations) field are generated which then propagate to the third lens. We require the output fluence maps to have spatial resolution $\Delta x_2 = 4.4\mu m$, which corresponds to the CCD pixel size. For the propagation from the third lens to the image plane, we use the Fresnel two-step propagator [61] for different step size requirements at the source and observation plane as per shown in Fig. 5.9. The source and observation plane side lengths are $L_1 = M\Delta x_1$, $L_2 = M\Delta x_2$, where $M = 2^{15}$ is the number of samples, Δx_1 and Δx_2 are the respective sampling intervals. The field at observation plane can be calculated as:

$$U_2(x_2, y_2) =$$

$$\frac{L_2}{L_1} e^{jkz} e^{j\frac{k\,(L_1-L_2)}{2z\ L_2}[-({x_2}^2+{y_2}^2)]} F^{-1}\left\{ e^{j\pi\lambda z\frac{L_1}{L_2}[-({f_X}_1{}^2+{f_Y}_1{}^2)]} F\left\{ U_1(x_1, y_1) e^{j\frac{k\,(L_1-L_2)}{2z\ L_1}[-({x_1}^2+{y_1}^2)]} \right\} \right\}, \tag{5.7}$$

where (f_{x_1}, f_{y_1}) being the spatial frequency coordinates corresponding to the spatial coordinates (x_1, y_1).

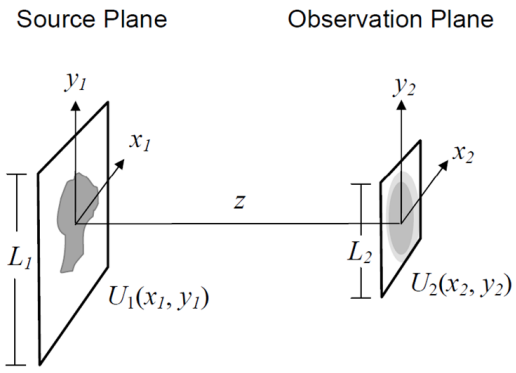

Figure 5.9 Fresnel two-step propagation geometry. Figure credit [61].

The ODWS model outputs are the three fluence maps F_0, F_x, F_y corresponding to the uniform transmission, horizontal, and vertical gradient filter transmission respectively. After propagation, the beam is demagnified by a factor of ~11 and is of size ~0.8 mm. It corresponds to 192×192 pixels which are cut out for further analysis.

5.5 Training and testing data generation, and results on simulation data

The strategy for generating the input wavefront data set is to provide sufficient wavefront variability while ensuring the wavefront slopes to be within the dynamic range of the ODWS, primarily limited by the proof-of-concept filter width of 2 cm. Ten thousand data sets were created for each of the three filters as shown in Fig. 5.3. Figure 5.10 shows the statistics of the 10,000-wavefront set generated for the training and testing purpose. The peak-to-valley of the input wavefronts is between 0 and 10 waves and the maximum slope for each wavefront is between -4.6 and 4.6 waves/mm.

The wavefront data consists of three sets. Set 1 is created with randomized individual fringe Zernike polynomial (Z_i) and corresponding coefficients (c_i), $\varphi(x, y) = \sum_{i=4}^{37} c_i Z_i(x, y)$. Set 2 is made of combined fringe Zernike polynomials with coefficients chosen from unique

normal distributions. Set 3 is the combinations of different Zernike wavefronts and random patterns.

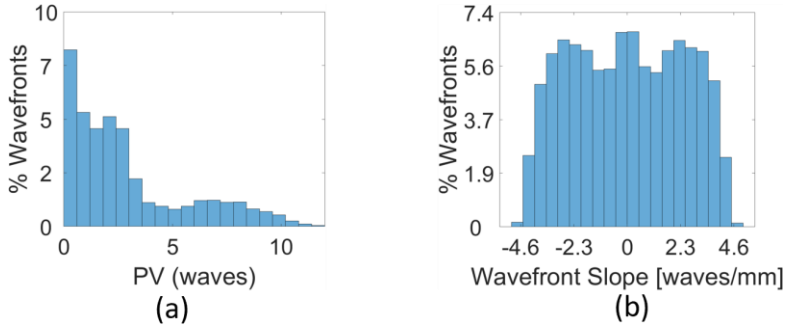

Figure 5.10 (a) Peak-to-valley (PV) and (b) wavefront slopes of the wavefront dataset.

A subset of the wavefront and fluence data (9,000 out of 10,000 pairs), obtained through beam propagation, was used to train the CNN. The remaining 1,000 pairs were used as test data to evaluate the CNN. Figures 5.11 (a), (b) and (c) show the histograms of the residual wavefronts which are the differences between the ground-truth and the CNN predicted wavefronts. The analytically reconstructed wavefronts (Eq. (4.14)) were compared to the ground truth to validate the beam propagation model, confirming a retrieval accuracy of less than 0.1λ RMS residual (Fig. 5.11 d). Figure 5.11 further shows that CNN-based wavefront retrievals for the two nonlinear filters have the same level of accuracy as the linear filter, with residual wavefronts being less than 0.1λ (RMS). The CNN is capable of retrieving the wavefront using a nonlinear filter, for which there is no analytical solution. The distribution of the residual wavefront (RMS) obtained from the analytic equation peaks at 0.05λ while this peak is lower than 0.02λ for the CNN reductions for all three filters. This demonstrates that the CNN is statistically a more robust method for ODWS wavefront reconstruction, in comparison to the analytic equation. This is because the analytic solution assumes slow varying amplitude and the accuracy also depends on the robustness of wavefront-slope integration [35, 62].

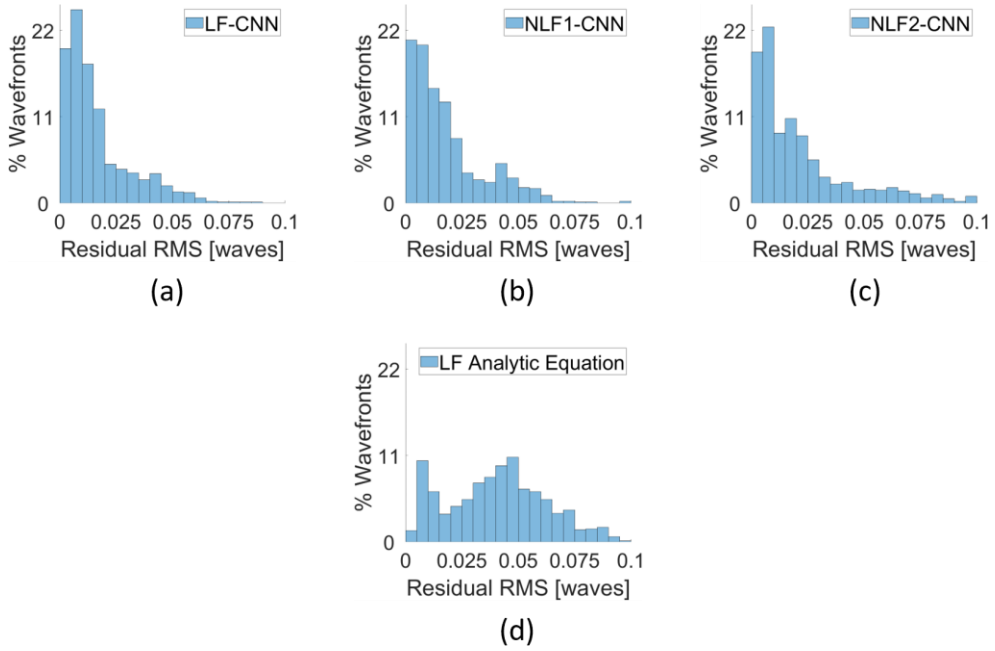

Figure 5.11 Residual RMS of the 1000 test wavefronts for (a) the linear filter LF, (b) the nonlinear filter NLF1, and (c) the nonlinear filter NLF2, retrieved by the CNN and (d) by the analytic equation.

5.6 Experimental setup for training and testing the developed CNN

We built an experimental setup as shown in Fig 5.12. This experiment uses the Spatial Light Modulator (SLM) that serves as a reconfigurable-CGH-based null in a freeform metrology system, developed by Romita et al. at the University of Rochester [9]. An SLM (Meadowlark Optics) with 1152×1920 pixels and a pitch of 9.2 μm is used to generate the training and test wavefronts. A laser beam (632.8-nm wavelength) from a Fizeau interferometer is down collimated by the 2:1 beam expander so that the beam size fits the short axis of the SLM. The circularly polarized light emitted by the interferometer is first converted to linear polarization by a quarter wave plate. Then a half wave plate is used to rotate the axis of the linear polarization to align it parallel to the extraordinary axis of the SLM. This polarization alignment is critical for minimizing unmodulated light. The SLM modulates the wavefront (9.73-mm pupil diameter) of the linearly polarized laser beam using birefringence. A tilt carrier (~200 waves) is added to the generated wavefront to separate the diffraction orders. The +1 order is used in the system. A standalone ODWS was built

using the nominal specifications as per Table 5.2 and was validated using the phase plates described in Chapter 4. The reflected +1 order from the SLM is aligned to the ODWS arm through the beam splitter. The ODWS camera is positioned at the imaging plane of the SLM; containing sharp images of crosshairs displayed on the SLM during alignment.

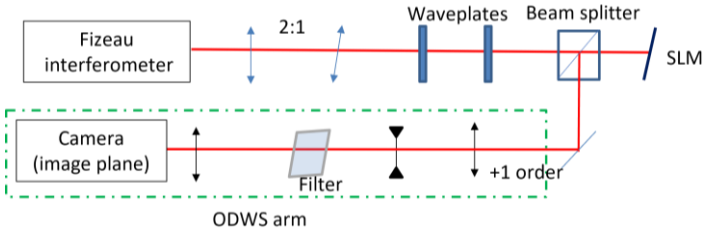

Figure 5.12 Schematic of the experimental ODWS setup to collect training and testing data.

We automated the SLM-based wavefront generation and the fluence data acquisition. The control loop for the SLM-ODWS is implemented in a Python program. The SLM is configured to generate a wavefront using a software operation in Matlab. The ODWS camera is then controlled to capture the fluence with certain exposure time. The exposure time is in the order of milli-seconds and is automatically scaled during the acquisition to keep the maximum pixel counts to 240 for the 8-bit camera.

We measured the SLM-ODWS path reference wavefront, $W_{sys,OD}$ using ODWS with a linear filter when no phase modulation is added on SLM except for the tilt. We then add $-W_{sys,OD}$ to the SLM to compensate the reference wavefront. Figure 5.13(a) shows the ODWS path system wavefront $W_{sys,OD}$, The astigmatic nature is due to the tilted lens after the Fizeau interferometer which avoids the reflected beam back to the interferometer. Figure 5.13(b) shows the ODWS reference wavefront when the compensating phase is added on SLM. The ODWS reference after the compensation is close to the diffraction limit.

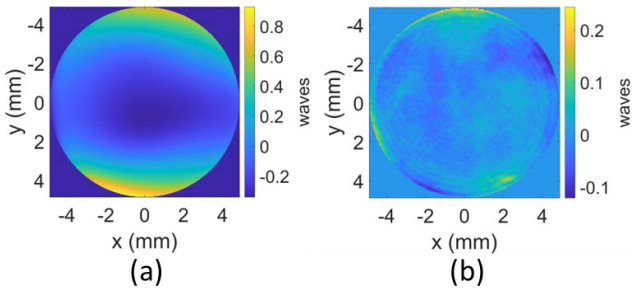

Figure 5.13 (a) ODWS path system wavefront, RMS 0.26λ, PV 1.27λ, (b) ODWS reference wavefront after compensated by the SLM, RMS 0.03λ, PV 0.36λ.

5.7 CNN performance on experimental data

The CNN training was further carried out experimentally, using the setup shown in Fig. 5.12. The same set of wavefronts used for theoretical training were created using the SLM, and the corresponding fluences were captured at the image plane of the ODWS. The CNN was then trained using pairs of fluence ratios (for filters in Fig. 5.3) and wavefronts for the 9,000-training data. To make a fair assessment of linear or nonlinear filter based retrieval, ground truth wavefronts in training should be used independent of ODWS, for example, calibrating the SLM or measuring with another wavefront sensor. In this work, we used the input command wavefront to the SLM as the ground truth. Figure 5.14 shows the training as the loss (averaged RMS wavefront residual across data) progression with epochs.

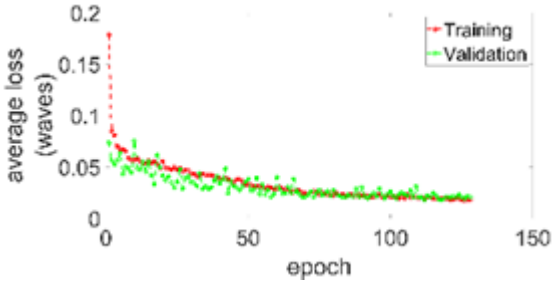

Figure 5.14 Loss progression showing CNN training.

The experimentally trained CNN was evaluated on the 1,000 test wavefronts. For the result shown in Fig. 5.15, residual wavefronts have less than 0.02λ RMS for 38.4%, 38.6% and 42.3% of the test wavefronts for the linear and the two nonlinear filters, respectively. It shows overall

nonlinear filter performs better, meaning that, given same dynamic range, the nonlinear filter retrieval appears to be more sensitive. The CNN reconstruction using all three filters [Figs. 5.15(a)-5.15(c)] has lower residual wavefront RMS than the linear analytic reconstruction [Fig. 5.15(d)], following the simulated data trend shown in Fig. 5.11. This provides an experimental validation that a CNN can reduce ODWS fluence data with better accuracy than the analytical method. Experimentally, the limitation on accuracy is due to detector noise, misalignments, and inaccuracy in amplitude transmission with spatial dithered binary pixelated filters. The CNN is much more robust to these system imperfections as compared to the analytic equation. Compared to the CNN retrievals [Figs. 5.11(a)-5.11(c) and Figs. 5.15(a)-5.15(c)], the accuracy deterioration in the experimental condition is much higher for the analytic solution [Fig. 5.11(d) and Fig. 5.15(d)].

We further plot the relative error, residual RMS/Ground Truth (GT) RMS vs. the Ground Truth RMS in Fig. 5.16. We observe a progressive increase in measurement accuracy for the wavefronts with low wavefront strengths, corresponding to low spatial frequency, as filters with larger center slopes are used for the LF, NLF1 and NLF2, respectively. This demonstrates improved measurement sensitivity, without sacrificing the dynamic range by using the nonlinear filter profiles. It demonstrates that CNN can retrieve wavefronts accurately across the dynamic range and the nonlinear filters show more sensitivity.

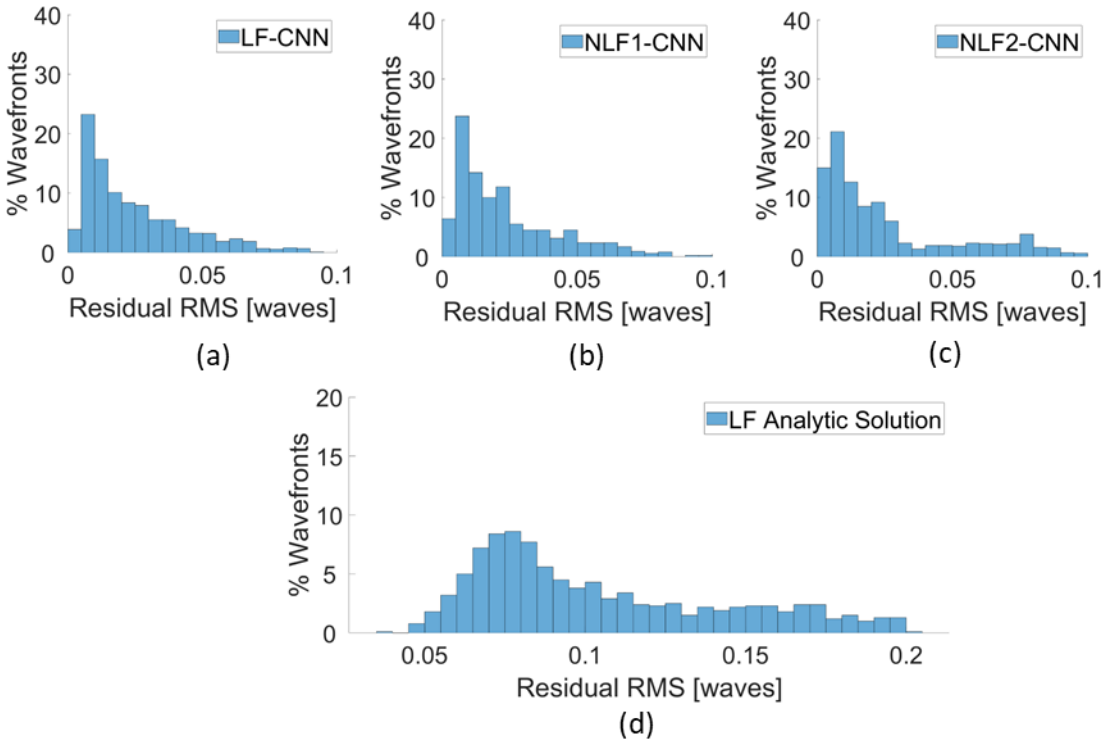

Figure 5.15 Residual RMS of the 1000 test wavefronts experimentally measured through (a) the linear filter LF, (b) the nonlinear filter NLF1, and (c) the nonlinear filter NLF2, retrieved by the CNN and (d) by the analytic equation.

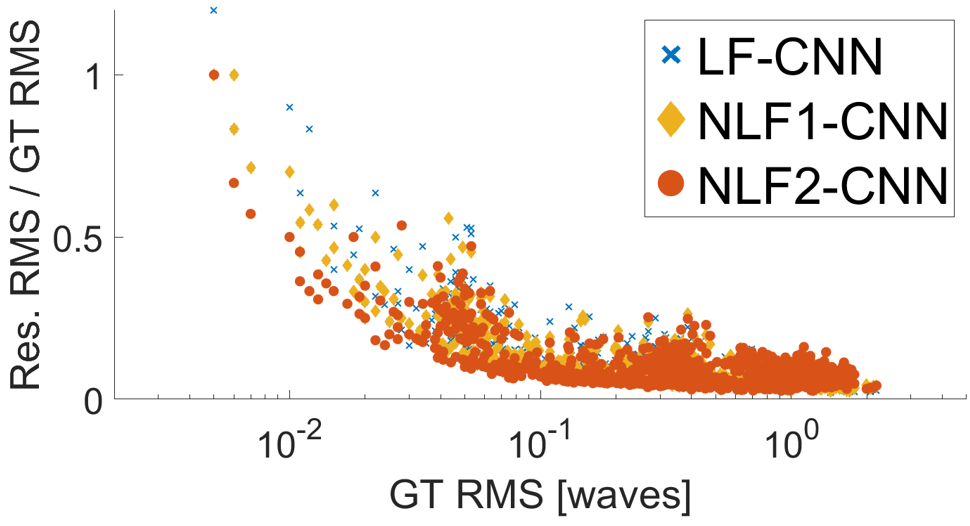

Figure 5.16 Comparison of the relative CNN reconstruction error of the 1,000 test wavefronts.

This is an experimental validation that a CNN can reduce ODWS fluence data, produced by binary pixelated filters with generalized transmission profiles, to wavefronts accurately. The local transmission slopes of the nonlinear filter shapes can be designed to tune measurement sensitivity to meet metrology requirements.

We tested the robustness of the CNN in predicting random phase profiles, which cannot be described in terms of basis functions such as Zernike coefficients or deformable mirror modes. The wavefront shown in Fig. 5.17(a) is a random phase profile which was generated using a technique in which the Fourier spectrum of a random number matrix is multiplied with a Gaussian filter [63]. The complexity and the slope of the wavefront is adjusted by setting the standard deviation of the Gaussian filter. The CNN prediction from the ODWS fluence images and its difference with the ground truth are plotted in Figs. 5.17(b) and 5.17(c) respectively. This result indicates that the CNN architecture is capable of predicting the complex shapes that can be present in astronomical imaging and biological samples.

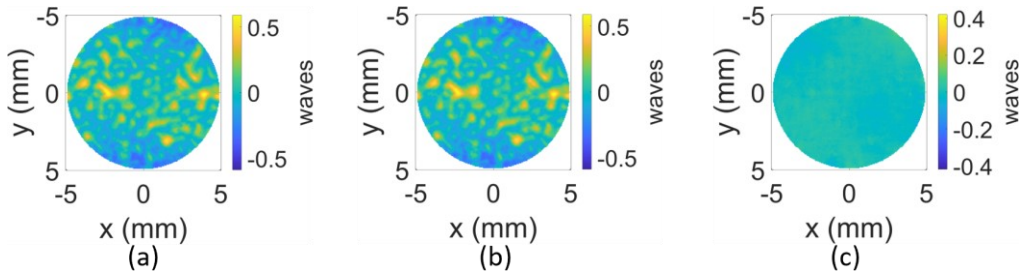

Figure 5.17 CNN performance on a random pattern wavefront: (a) Analytic reconstruction (PV: 1.16λ, RMS: 0.15λ); (b) CNN prediction (PV: 1.14λ, RMS: 0.15λ); (c) Residual (RMS: 0.02λ).

Chapter 6 Subaperture stitching of wavefronts

There are potentially three benefits of subaperture stitching with ODWS. First, when measuring optical phase components of aperture larger than the ODWS diagnostic aperture, several measurements over the diagnostic aperture can be stitched to obtain the full aperture wavefront. The second potential benefit of subaperture stitching is to avoid the effect of any local surface tilts in a freeform surface which otherwise leads to extended far-field and thus reduces the dynamic range. This local tilt can be cancelled by aligning the subaperture containing the local surface tilt such that it is perpendicular to the beam path. Then the subapertures can be stitched and local tilts will be computed. The third potential benefit of stitching is that the accuracy of ODWS increases with decreasing aperture size for a fixed ODWS filter pixel size. This is because of more sampling points in a larger focal spot in the filter plane. Then the accurately measured subapertures can be stitched to characterize over full aperture size.

6.1 Simulation of stitching algorithm

The subaperture stitching methods have been investigated for stitching interferometry [64,65], null based interferometry [66] and scanning Shack Hartman wavefront sensor [67] in literature. The algorithms are based on the least square solution to determine misalignments of subapertures and include higher complexities specific to the measurement method. In this work, we consider ideal wavefront measurement condition and implement the algorithm described in [65].

The subaperture measurements can be performed by moving the sample on a translation stage in lateral directions to cover the full aperture. The translation stage can move laterally with incremental movement of less than 1μm, which is less than the lateral resolution of ODWS (~50 μm). Therefore, the error of lateral position can be ignored. Error from tip/tilt and displacement

along optical axis (piston) are of concern and need to be corrected during stitching. To correct these errors, ODWS measurements will be conducted in a way that the adjacent subapertures have overlapping areas. Then the subapertures will be stitched such that difference in phase distribution in the common area is minimum.

Let there be two subaperture areas A and A' with coordinates (x, y) and (x', y') respectively as shown in Fig. 6.1. If (x_0, y_0) is the shift of origin of between the two subapertures, then $x = x' + x_0$ and $y = y' + y_0$. Let the wavefront of the two subapertures be $\varphi(x, y)$ and $\varphi'(x - x_0, y - y_0)$, respectively.

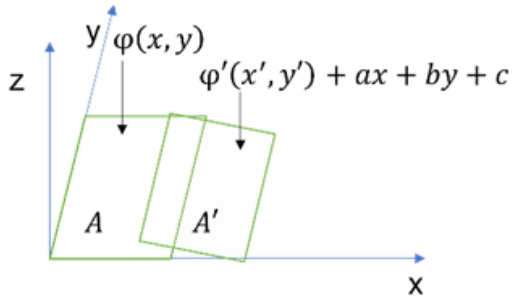

Figure 6.1 Two subapertures A and A' having overlapping are measured.

Considering wavefront of subaperture A as a standard measurement, the relation between φ and φ' in the overlapping area of A and A' satisfies the following,

$$\varphi(x, y) = \varphi'(x - x_0, y - y_0) + ax + by + c \tag{6.1}$$

where, a and b are coefficients of tip and tilt errors, and c is the piston error in measuring φ'. These errors can be minimized by finding the coefficients of tip/tilt/piston and correcting the wavefront accordingly. The coefficients can be found by minimizing the sum of the squared difference of the phase distributions in the overlapping area of two subapertures:

$$\text{Minimize:} \qquad \sum\{\varphi(x, y) - [\varphi'(x - x_0, y - y_0) + ax + by + c]\}^2 \tag{6.2}$$

Differentiating Eq. (6.2) with respect to (x, y), the following relation can be obtained:

$$\begin{bmatrix} \sum x(\varphi-\varphi') \\ \sum y(\varphi-\varphi') \\ \sum(\varphi-\varphi') \end{bmatrix} = \begin{bmatrix} \sum x^2 & \sum xy & \sum x \\ \sum xy & \sum y^2 & \sum y \\ \sum x & \sum y & n \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix}, \tag{6.3}$$

where n is the number of sampling points in the overlapping area. From Eq. (6.3), the misalignment coefficients can be obtained as:

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \sum x^2 & \sum xy & \sum x \\ \sum xy & \sum y^2 & \sum y \\ \sum x & \sum y & n \end{bmatrix}^{-1} \begin{bmatrix} \sum x(\varphi-\varphi') \\ \sum y(\varphi-\varphi') \\ \sum(\varphi-\varphi') \end{bmatrix} \tag{6.4}$$

In the simulation in MATLAB, a full aperture freeform wavefront of dimension 12 mm × 12 mm as shown in Fig. 6.2(a) was divided into 2 x 2 square subaperture arrangement as shown in Fig. 6.2(b), not to scale.

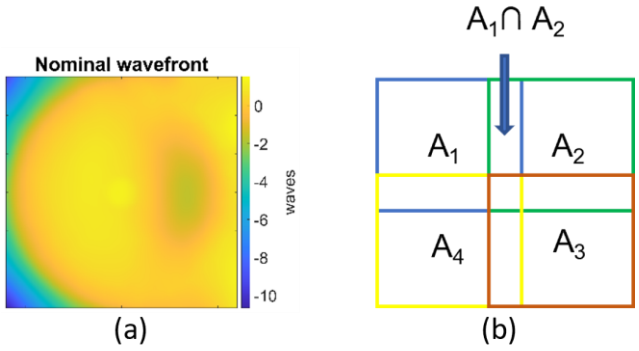

Figure 6.2 (a) Nominal full aperture freeform wavefront with full area of 12 mm x 12mm: Peak-to-valley (PV) 12.2λ, Root-mean-square (RMS) 1.65λ, λ=633nm. (b) Arrangement of 2 x 2 measurement areas of overlapping subapertures A_1, A_2, A_3 and A_4, (not to scale).

The wavefront was divided such that the overlapping area between two subapertures is 50% of a subaperture. The subaperture A_1 was taken as standard. The nominal subapertures are shown in Fig. 6.3. Alignment errors i.e., piston, tip and tilt of different amplitudes were added to subapertures A_2, A_3 and A_4. The misaligned subapertures are shown in Fig. 6.4. Alignment errors added exceed the nominal wavefront as seen from the scale in color bar (PV of A_2: ~50λ in Fig. 6.4 vs ~3λ in Fig. 6.3).

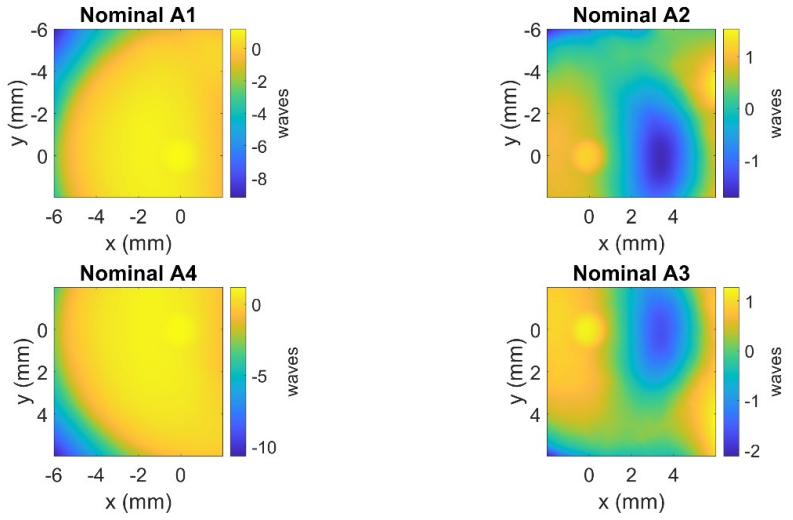

Figure 6.3 Nominal subapertures extracted from full aperture wavefront of Fig. 6.2(a). Area overlap between two subapertures is 50%.

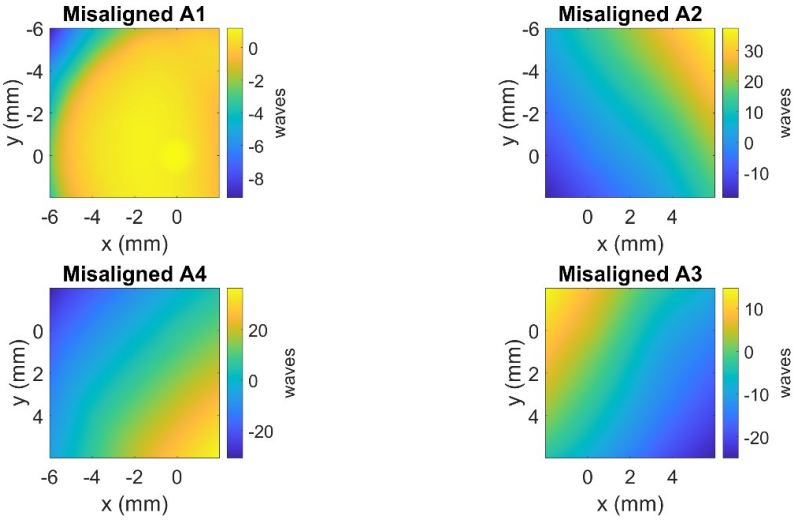

Figure 6.4 Subapertures after piston, tip and tilt misalignments added to A_2, A_3 and A_4. Alignment errors added exceed the nominal wavefront as seen from the scale in color bar.

For each overlapping region, the data of phase distribution, spatial coordinates and number of sampling points were extracted and used to solve for the coefficients of alignment errors of each misaligned subaperture using Eq. (6.4). Then the subapertures were aligned using the retrieved coefficients of misalignment and stitched to get full aperture wavefront map. We followed a sequential stitching of two subapertures by stitching the subapertures pairs in the order (A_1, A_2), (A_2, A_3) and (A_3, A_4). A_2 can be aligned with respect to A_1(standard) and then A_2 becomes the

standard for A_3. The result in Fig. 6.5 demonstrates that the subapertures are stitched accurately and the exact wavefront as the original one was produced. This process can be extended to subaperture arrangement of 3 × 3 and so on.

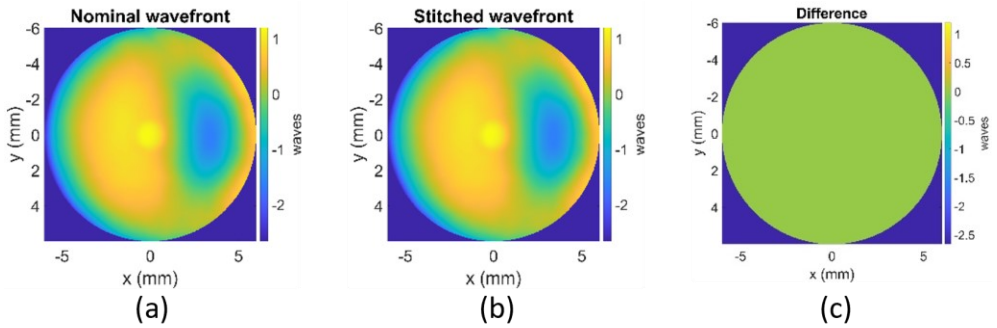

Figure 6.5 Comparison between nominal and stitched wavefront inside pupil of 12 mm diameter: (a) Nominal wavefront PV 3.86λ, RMS 0.71λ (b) Stitched wavefront PV, 3.86λ RMS 0.71λ (c) Difference PV 0λ, RMS 0λ.

Chapter 7 Conclusion and future direction

We have demonstrated the first high-performance Optical Differentiation Wavefront Sensor (ODWS) based on spatially dithered binary pixelated filters towards freeform metrology. Its performance in freeform phase plate measurement is compared to that of a commercial scanning low-coherence interferometer, and shows excellent agreement. The accuracy in comparison and precision have been determined to be ~ λ/10 (RMS) and ~ λ/70 (RMS), respectively. We experimentally demonstrated that the dynamic range can be increased via tailoring the filter width or focal length of the imaging system. Though coordinate-based scanning techniques can measure high surface slopes, ODWS has the advantage of short measurement times. Coordinate-based measurements typically have micron-level precision. Our ODWS demonstration has shown precision of λ/70 that is comparable to nanometer level precision in interferometry. Simulations that include the photodetection noise and the far-field filter nonlinearity demonstrate the impact of these parameters and the general robustness of the ODWS.

We have formulated the concept of a compact ODWS based on a telephoto configuration and experimentally demonstrated wavefront retrieval with such diagnostic based on spatially dithered binary pixelated filters. The ODWS measurements of phase plates are consistent with those from a commercial scanning low-coherence interferometer. The total system length is five times shorter than a $4f$ ODWS system of identical effective focal length, yet allowing for accurate metrology of free-form optics. We have shown experimentally that the demonstrated ODWS has an alignment tolerance, rendering accurate measurements, that can easily be met in practical conditions. This novel implementation of an ODWS with a telephoto-lens expands the domain of applicability for this technique.

We demonstrated that CNN can be a robust wavefront reconstructor when ODWS is employed with pixelated linear and nonlinear filters, allowing for high dynamic range, high sensitivity and high-resolution wavefront sensing. We developed a numerical beam propagation model to facilitate the theoretical training of a CNN. The CNN architecture reconstructs the wavefront in a zonal manner from the ODWS fluences. It was also found that this architecture retrieves wavefront with better accuracy than that of a CNN producing wavefront modal coefficients. We further demonstrated an SLM wavefront generator with ODWS to perform experimental training and testing. The CNN was separately trained from scratch with simulation and experimental data. For both cases, a typical residual of less than 0.05λ was achieved. We achieved improved sensitivity within a fixed dynamic range, overcoming the tradeoff between dynamic range and sensitivity.

We tested a subaperture stitching algorithm to sense a wavefront that is larger than the ODWS diagnostic aperture. The algorithm recovers the full wavefront exactly using the overlapping measurement area of the misaligned subaperture measurements.

The main advantages of the ODWS are its scalable dynamic range and the intrinsic high resolution compared to other wavefront sensors. As an example, a commercial Shack Hartman sensor WFS20-5C from Thorlabs has microlenses of diameter 150-μm, focal length of 4.1 mm and wavefront diameter of 5.4-mm. When measuring a beam/object with 20-mm diameter, considering the magnification factor of 3.7, the expected spatial resolution and dynamic range for wavefront slope will be 555.6-μm and ±7.6 waves/mm, respectively. For comparison with the implemented telephoto ODWS system for measuring the same 20-mm beam, the expected ODWS spatial resolution and dynamic range are 50.2-μm and ±11.5 waves/mm, respectively when the filter width is increased to 5-cm. The spatial resolution is less than one tenth and the dynamic range is 1.5

times that of the expected values from the Shack Hartman sensor. If the beam size is reduced, the ODWS dynamic range will be further increased proportionally because the effective focal length can be reduced by the same factor. Furthermore, commercial lithography allows for scaling the filter width to the level of tens of centimeters, if needed. For a filter size of 10-cm, the expected spatial resolution and dynamic range are 50.2-μm and ±23 waves/mm, which is significantly higher than what a Shack Hartman sensor can offer.

In the future perspective, the calibration of the image sensor can be done to account for any nonlinearity and nonuniform response. Improvements in the wavefront slope integration algorithm can be further investigated. For CNN based retrieval, the experimental efficiency can be further improved by transferring the ODWS-simulation-trained CNN parameters, such as weights for fine tuning the experimental data. Further, the experimental ground truth can be characterized with another wavefront sensor. The subaperture stitching algorithm can be further investigated by propagating and retrieving the subaperture wavefronts through ODWS model. In simulation, more uncertainty such as noise and other misalignment errors can be included to study impact on stitching accuracy. Then the algorithm can be tested in experimental condition. The scalability of ODWS for more complex wavefronts that require further high dynamic range and high resolution measurement can be investigated further by implementing filters of larger width and focusing lens with shorter focal length and optimizing experimental system parameters.

Bibliography

1. J. P. Rolland, M. A. Davies, T. J. Suleski, C. Evans, A. Bauer, J. C. Lambropoulos, and K. Falaggis, "Freeform optics for imaging," Optica 8(2), 161 (2021).
2. S. Wills, "Freeform Optics: Notes from the Revolution," Optics and Photonics News 28(7), 34 (2017).
3. A. M. Herkommer, "Advances in the design of freeform systems for imaging and illumination applications," Journal of Optics 43(4), 261–268 (2014).
4. K. Fuerschbach, J. P. Rolland, and K. P. Thompson, "A new family of optical systems employing φ-polynomial surfaces," Opt. Express 19(22), 21919–21928 (2011).
5. F. Z. Fang, X. D. Zhang, A. Weckenmann, G. X. Zhang, and C. Evans, "Manufacturing and measurement of freeform optics," CIRP Annals - Manufacturing Technology 62(2), 823–846 (2013).
6. E. Savio, L. de Chiffre, and R. Schmitt, "Metrology of freeform shaped parts," CIRP Annals 56(2), 810–835 (2007).
7. D. Malacara, Optical shop testing (John Wiley & Sons, 2007), Vol. 59.
8. L. Huang, H. Choi, W. Zhao, L. R. Graves, and D. W. Kim, "Adaptive interferometric null testing for unknown freeform optics metrology," Opt. Lett. 41, 5539-5542 (2016).
9. R. Chaudhuri, A. Wansha, R. Porras-Aguilar, and J. P. Rolland, "Implementation of a null test for freeform optics using a high-definition spatial light modulator," Opt. Express 30(24), 43938–43960 (2022).
10. B. C. Platt and R. Shack, "History and principles of Shack-Hartmann wavefront sensing," J. refractive surgery 17, 573–577 (1995).

11. M. Rocktäschel and H. J. Tiziani, "Limitations of the Shack–Hartmann sensor for testing optical aspherics," Optics & Laser Technology 34, 631-637 (2002).
12. N. Lindlein, J. Pfund, and J. Schwider, "Algorithm for expanding the dynamic range of a Shack-Hartmann sensor by using a spatial light modulator," Optical Engineering 40(2001).
13. M. Aftab, H. Choi, R. Liang, and D. W. Kim, "Adaptive Shack-Hartmann wavefront sensor accommodating large wavefront variations," Opt. Express 26, 34428-34441 (2018).
14. Y. Saita, H. Shinto, and T. Nomura, "Holographic Shack–Hartmann wavefront sensor based on the correlation peak displacement detection method for wavefront sensing with large dynamic range," Optica 2, 411-415 (2015).
15. J. Lee, R. V. Shack, and M. R. Descour, "Sorting method to extend the dynamic range of the Shack–Hartmann wave-front sensor," Appl. Opt. 44, 4838-4845 (2005).
16. J. Pfund, N. Lindlein, and J. Schwider, "Dynamic range expansion of a Shack–Hartmann sensor by use of a modified unwrapping algorithm," Opt. Lett. 23, 995-997 (1998).
17. D. R. Burada, K. K. Pant, V. Mishra, M. Bichra, G. S. Khan, S. Sinzinger, and C. Shakher, "Development of a metrology technique suitable for in situ measurement and corrective manufacturing of freeform optics," Advanced Optical Technologies 8, 203-215 (2019).
18. K. K. Pant, D. R. Burada, M. Bichra, M. P. Singh, A. Ghosh, G. S. Khan, S. Sinzinger, and C. Shakher, "Subaperture stitching for measurement of freeform wavefront," Appl. Opt. 54, 10022-10028 (2015).
19. G. Baer, J. Schindler, C. Pruss, J. Siepmann, and W. Osten, "Calibration of a non-null test interferometer for the measurement of aspheres and free-form surfaces," Optics express 22, 31200-31211 (2014).

20. D. Xu, R. Chaudhuri, and J. P. Rolland, "Telecentric broadband objective lenses for optical coherence tomography (OCT) in the context of low uncertainty metrology of freeform optical components: from design to testing for wavefront and telecentricity," Optics express 27, 6184-6200 (2019).
21. D. Xu, J. Yao, N. Zhao, and J. P. Rolland, "Scanning Customized Swept-source Optical Coherence Tomography (SS-OCT) for the Metrology of Freeform Optical Surfaces," in Frontiers in Optics 2016, OSA Technical Digest (online) (Optical Society of America, 2016), FW5H.6.
22. D. W. Diehl, C. J. Ditchman, and C. T. Cotton, "Low-coherence surface metrology using a multiple-beam optical probe," in Window and Dome Technologies and Materials XI, (International Society for Optics and Photonics, 2009), 73020T.
23. L. Graves, H. Quach, H. Choi, and D. Kim, "Infinite deflectometry enabling 2π-steradian measurement range," Optics express 27, 7602-7615 (2019).
24. C. Faber, E. Olesch, R. Krobot, and G. Häusler, "Deflectometry challenges interferometry: the competition gets tougher!," in Interferometry XVI: Techniques and Analysis, (International Society for Optics and Photonics, 2012), 84930R.
25. K. Medicus, S. DeFisher, M. Bauza, and P. Dumas, Round-robin measurements of toroidal window, SPIE Optifab (SPIE, 2013), Vol. 8884.
26. R. A. Sprague and B. J. Thompson, “Quantitative visualization of large variation phase objects,” Appl. Opt.11(7), 1469–1479 (1972).
27. J. C. Bortz, “Wave-front sensing by optical phase differentiation,” J. Opt. Soc. Am. A 1(1), 35–39 (1984).

28. J. E. Oti, V. F. Canales, and M. P. Cagigal, “Analysis of the signal-to-noise ratio in the optical differentiation wavefront sensor,” Opt. Express11(21), 2783–2790 (2003).
29. F. Hénault, “Wavefront sensor based on varying transmission filters: theory and expected performance,” J. Mod. Opt.52(14), 1917–1931 (2005).
30. J. Oti, V. Canales, and M. Cagigal, “Improvements on the optical differentiation wavefront sensor,” Mon. Not. R.Astron. Soc.360(4), 1448–1454 (2005).
31. M. P. Cagigal and P. J. Valle, “x–y curvature wavefront sensor,” Opt. Lett.40(8), 1655–1658 (2015).
32. T. Szoplik, V. Climent, E. Tajahuerce, J. Lancis, and M. Fernández-Alonso, “Phase-change visualization in two-dimensional phase objects with a semi derivative real filter,” Appl. Opt.37(23), 5472–5478 (1998).
33. A. Sagan, T. J. Antosiewicz, and T. Szoplik, “Three filters for visualization of phase objects with large variations of phase gradients,” Appl. Opt.48(6), 1143–1152 (2009).
34. J. Qiao, Z. Mulhollan, and C. Dorrer, “Optical differentiation wavefront sensing with binary pixelated transmission filters,” Opt. Express24(9), 9266–9279 (2016).
35. J. Qiao, A. Travinsky, G. Ding, and C. Dorrer, Optical differentiation wavefront sensor based on binary pixelated transmission filters, SPIE LASE (SPIE, 2015), Vol. 9356.
36. W. H. Southwell, “Wave-front estimation from wave-front slope measurements,” J. Opt. Soc. Am.70(8), 998–1006(1980).
37. D. Schmidt and O. von der Lühe, “Optical wavefront differentiation: wavefront sensing for solar adaptive optics based on a LCD,” in Adaptive Optics for Laser Systems and Other Applications, (International Society for Optics and Photonics, 2007), 658408.

38. H. Furuhashi, J. V. Mayorga, Y. Uchida, and A. Kono, “Phase measurement of optical wavefront by an SLM differentiation filter,” Laser 15, 10 (2009).
39. S. Haffert, “Generalised optical differentiation wavefront sensor: a sensitive high dynamic range wavefront sensor,” Opt. Express 24(17), 18986–19007 (2016).
40. E. Gendron, M. Brangier, G. Chenegros, F. Vidal, Z. Hubert, G. Rousset, and F. Pouplard, “A new sensor for laser tomography on ELTs,” in 1st AO4ELT conference-Adaptive Optics for Extremely Large Telescopes, (EDP Sciences, 2010), 05003.
41. R. Landman and S. Haffert, “Nonlinear wavefront reconstruction with convolutional neural networks for fourier-based wavefront sensors,” Opt. Express 28(11), 16644–16657 (2020).
42. H. Furuhashi, K. Matsuda, and C. P. Grover, “Visualization of phase objects by use of a differentiation filter,” Appl. Opt. 42(2), 218–226 (2003).
43. C. Dorrer and J. Qiao, “Direct binary search for improved coherent beam shaping and optical differentiation wavefront sensing,” Appl. Opt. 57, 8557–8565 (2018).
44. C. Dorrer and J. D. Zuegel, “Design and analysis of binary beam shapers using error diffusion,” J. Opt. Soc. Am. B 24(6), 1268–1275 (2007).
45. J. Tierson, E. Fess, and G. Matthews, “Developments in precision asphere manufacturing,” in Optifab 2015, (International Society for Optics and Photonics, 2015), 96330H.
46. DeFisher, S., Matthews, G., Fess, E., “Validation of accuracy and repeatability of UltraSurf metrology on common optical shapes,” 96331M (2015).
47. Dufour M.L, Lamouche G., Detalle V., Gauthier B., Sammut P. “Low-Coherence Interferometry, an Advanced Technique for Optical Metrology in Industry”. Non-Destructive Testing & Condition Monitoring, 47(4):216-219, (2005).

48. J. Babington, “Alvarez lens systems: theory and applications,” in Optical Systems Design 2015: Optical Design and Engineering VI, (International Society for Optics and Photonics, 2015), 962615.
49. C. Dorrer and J. Hassett, “Model-based optimization of near-field binary-pixelated beam shapers,” Appl. Opt. 56(4), 806–815 (2017).
50. J. W. Goodman, “Introduction to Fourier Optics,” (W. H. Freeman, 2017).
51. R. W. Gerchberg, “A practical algorithm for the determination of phase from image and diffraction plane pictures,” Optik 35, 237–246 (1972).
52. J. R. Fienup, “Phase retrieval algorithms: a comparison,” Appl. Opt. 21, 2758–2769 (1982).
53. Y. Nishizaki, M. Valdivia, R. Horisaki, K. Kitaguchi, M. Saito, J. Tanida, and E. Vera, “Deep learning wavefront sensing,” Opt. Express 27(1), 240–251 (2019).
54. S. W. Paine and J. R. Fienup, “Machine learning for improved image-based wavefront sensing,” Opt. Lett. 43, 1235–1238 (2018).
55. R. Swanson, M. Lamb, C. Correia, S. Sivanandam, and K. Kutulakos, “Wavefront reconstruction and prediction with convolutional neural networks,” in Adaptive Optics Systems VI, vol. 10703 (International Society for Optics and Photonics, 2018), p. 107031F.
56. J. Dou, D. Wang, Q. Yu, M. Kong, L. Liu, X. Xu, and R. Liang, “Deep-learning-based deflectometry for freeform surface measurement,” Opt. Lett. 47(1), 78–81 (2022).
57. T. B. DuBose, D. F. Gardner, and A. T. Watnik, “Intensity-enhanced deep network wavefront reconstruction in Shack–Hartmann sensors,” Opt. Lett. 45(7), 1699–1702 (2020).
58. L. Hu, S. Hu, W. Gong, and K. Si, “Deep learning assisted Shack–Hartmann wavefront sensor for direct wavefront detection,” Opt. Lett. 45(13), 3741–3744 (2020).

59. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) (Springer Verlag, 2015), 9351, pp. 234–241.
60. A. Paszke, S. Gross, F. Massa, et al., "Pytorch: An imperative style, high-performance deep learning library", Adv. Neural Inf. Process. Syst. 32, 8026–8037 (2019).
61. D. G. Voelz, Computational Fourier Optics: A MATLAB Tutorial (SPIE, 2011).
62. B. Pathak and B. R. Boruah, "Improved wavefront reconstruction algorithm for Shack–Hartmann type wavefront sensors", J. Opt. 16, 055403 (2014).
63. S. Schott, J. Bertolotti, J.-F. Léger, L. Bourdieu, and S. Gigan, "Characterization of the angular memory effect of scattered light in biological tissues," Opt. Express 23(10), 13505–13516 (2015).
64. P. Murphy, G. Forbes, J. Fleig, P. Dumas, and M. Tricard, "Stitching interferometry: a flexible solution for surface metrology," Opt. Photonics News 14(5), 38–43 (2003).
65. M. Otsubo, K. Okada, and J. Tsujiuchi, "Measurement of large plane surface shapes by connecting small-aperture interferograms," Opt. Eng. 33, 608–613 (1994).
66. J. H. Burge, P. Su, and C. Zhao, "Optical metrology for very large convex aspheres," Proc. SPIE 7018, 701818 (2008).
67. K. Pant, D. Burada, M. Michra, M. Bichra, M. Singh, A. Ghosh, G. Khan, S. Sinzinger, and C. Shakher, "Subaperture stitching for measurement of freeform wavefront," Appl. Opt. 54(34), 10022–10028 (2015).