Estimating phase aberrations from intensity data

Theam Yong Chew and Richard G. Lane
Electrical and Computer Engineering,
University of Canterbury, Christchurch, New Zealand
{tyc20, rlane}@elec.canterbury.ac.nz

Abstract

Image processing has traditionally relied on improving images by estimation of the psf and then compensating the image using computer post processing techniques. In reality much of the distortion can be traced to phase aberrations caused by imperfections in the instrument optics. Modern technology means that these aberrations can be compensated, but the problem of estimating the aberration remains. In this paper we explain a new way of estimating the phase aberration from two defocused measurements of the instrument aperture.

Keywords:adaptive optics, image restoration, wavefront sensing

1 Introduction

The fundamental limit in the resolution of an optical system is inversely proportional to the diameter of the aperture of the optical system. However, imperfections in the optical system, or imaging through a turbulent media (for example, the earth's atmosphere), cause blurring of the images. In image processing, this is commonly modelled by a linear space invariant system, whereby a point-spread-function (PSF) may be defined. In optics, it is more useful to define the blurring in terms of phase errors in the aperture plane of the optical system, since these show how they can be corrected using deformable optics [1]. Figure 1 shows an example of the effect of wavefront aberrations on the PSF of an optical system.

Although aberrations of the wavefront phase are the most significant cause of distorted images, it is not possble with the current technology to measure the phase aberrations. Instead wavefront sensors work by inferring the phase indirectly from its effects on intensity measurements. The simplest way to transform phase distortions into measurable intensity distortions is to allow the wavefront to propagate as illustrated in figure 2. As is apparent from the figure, the further the wavefront is allowed to propagate the larger the measurable intensity fluctuations that are caused by the phase aberration. In practice, a lens (or mirror) is used to provide the same effective propagation distance within the confines of an instrument.

As shown in figure 2, increasing the propagation distance also increases the non-linearity of the relationship between the phase and the measured intensities. In the limit when the wavefront is propagated to infinity (or in the case of an instrument to the focal plane) the light propagation needs to be described using Fourier optics. However, provided the wavefront propagation

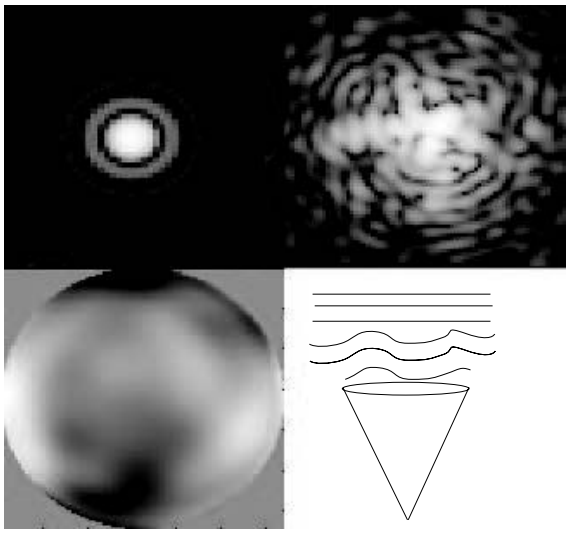

Figure 1: The diffraction limited PSF is shown on the upper left, and the aberrated PSF is shown to its right, with its corresponding wavefront aberration on the lower left. The lower right diagram illustrates the situation frequently encountered in astronomical imaging in the distortion of star light caused by atmospheric turbulence.

distance is limited, then geometric optics may be employed to describe the relationship between the phase and the magnitude.

1.1 Geometric optics

The key to geometric optics is the ability to assume that light propagates in a direction perpendicular to the wavefront. Figure 3 shows the basis of this model. For a given propagation distance, the light emanating from a point in the aperture is displaced in the image plane by a distance proportional to the slope of the wavefront.

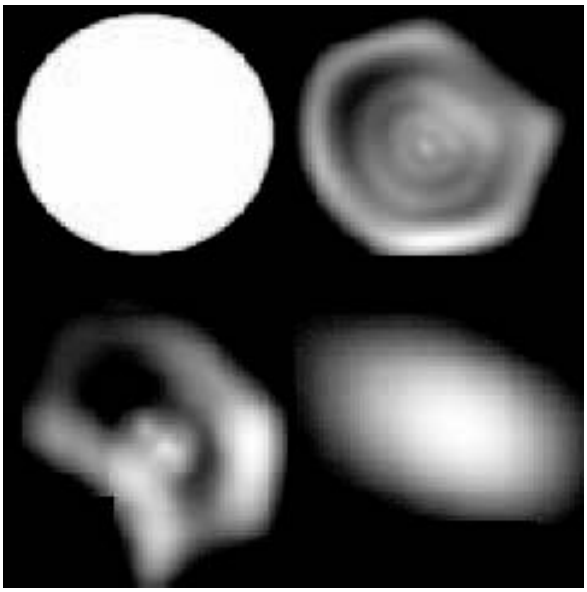

Figure 2: Simulated image showing intensity distributions of an aperture at (clockwise from top left) the aperture plane, and propagated to 30km, 300km and 70km. Note that the images have been rescaled for presentation purposes.

It is this linear relationship that is used as a basis for wavefront sensing.

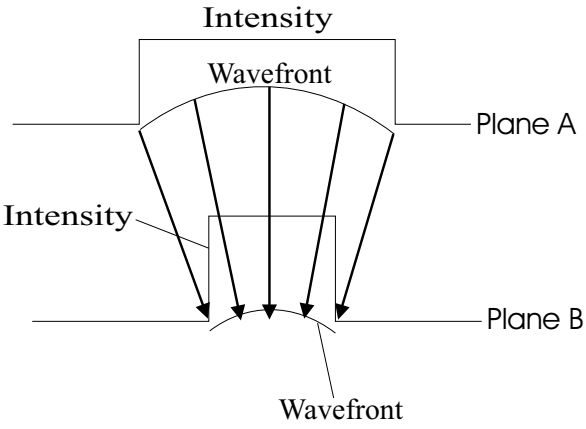

Figure 3: Geometric optics model for the propagation of light.

However, a problem arises when the light is allowed to propagate too large a distance, because in this case light rays from different parts of the aperture may cross over before reaching the measurement plane. In this case it is no longer possible to unambigously relate the light measured in the image to a location in the aperture. Not allowing the light to propagate to the focal plane is however at a cost in sensitivity as discussed in [2].

The simplest wavefront sensing device consists of two measurement planes each symmetrically distributed around the focal plane of the instrument. This data has traditionally been used to formulate an estimate of the aberration curvature [3], but we show how it is more appropriate to estimate the slope of the aberration. Section 2 compares these two approaches, and discusses how we can estimate the phase aberration from the slopes. Section 4 shows experimental results from this technique.

2 Interpretation of measurement data

A converging lens has the effect of causing light to concentrate evenly as the wavefront propagates. According to the geometric optics model the image taken at the measurement plane between the aperture and the focal plane (Plane A in figure 4) would thus be identical to the aperture, except that it is smaller and brighter.

Geometric optics also predicts that after having passed through a focal point the light would again start to spread out. At the second measurement plane on the other side of focus (Plane B in figure 4) one would expect to see another image of the aperture, except that this time it would both be enlarged and inverted from figure 4. In the absence of a phase aberration, the measurements at planes A and B planes would be identical.

The difference between estimating the curvature and the slopes can be seen from figure 4 which also shows the effect of a simple defocus in the aperture. The wavefront aberration can be written mathematically as

$$W(x,y) = \frac{k}{4}(x^2+y^2) \quad \forall x^2+y^2 < r^2 \tag{1}$$

where x and y are spatial coordinates in the plane transverse to the direction of propagation, k a constant representing the curvature, and r is the radius of the aperture.

The slope in the x and y directions is then

$$W_x(x,y) = \frac{k}{2}x, W_y(x,y) = \frac{k}{2}y \tag{2}$$

and the curvature is equal to

$$\nabla^2 W(x,y) = W_{xx}(x,y) + W_{yy}(x,y) = k \tag{3}$$

2.1 Estimating curvature

A direct curvature estimate is made by differencing two measurements in the two planes as shown in figure 5. A negative curvature at point P_1 in the figure causes the corresponding point at plane A to be brighter than B. The positive curvature at P_2 has the opposite effect. Essentially a local curvature in the wavefront will cause a dimming on one measurement plane and a brightening on the other.

Mathematically this is equal to

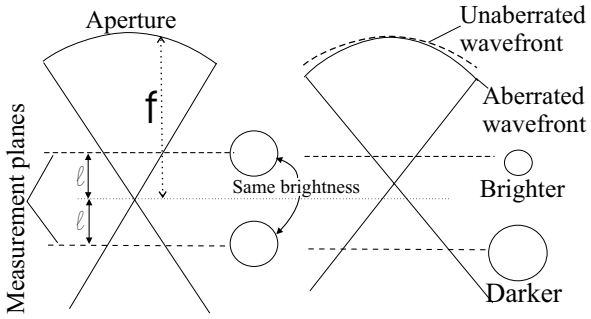

Figure 4: Comparison of an unaberrated wavefront (left) with a defocus (right). The "circles" represent intensity distribution in measurement planes coming out of the page.

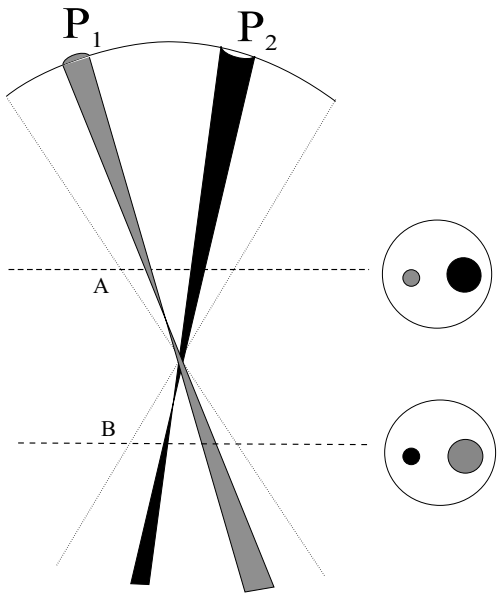

Figure 5: Detection of a hypothetical wavefront with local defocussing aberrations.

$$\nabla^2 W(\frac{f\mathbf{r}}{l}) = -\frac{f-l}{l}\left(\frac{I_A(\mathbf{r}) - I_B(-\mathbf{r})}{2I}\right) \quad (4)$$

$$\approx -\frac{f-l}{l}\left(\frac{I_A(\mathbf{r}) - I_B(-\mathbf{r})}{I_A(\mathbf{r}) + I_B(-\mathbf{r})}\right) \quad (5)$$

A complete analysis of this approach is beyond the scope of this paper, but can be found in[4]. The major deficiency in most analyses is that they neglect the finite nature of the instrument aperture, the effects of which are best shown graphically.

Figure 6 shows a one-dimensional defocus example. When the signals at the two measurement planes are differenced, the difference between intensities near the centre of the aperture indicates a constant curvature as expected, but at the edge the signal is incorrect. Integrating the signal two times to produce the wavefront produces a quantity that exactly matches neither the expected defocus nor the wavefront slope, as shown in figure 6.

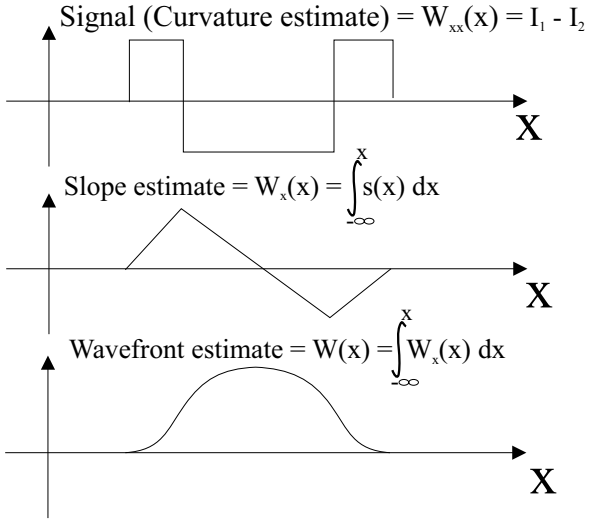

Figure 6: Estimation of a wavefront from curvature sensing data.

A further problem that occurs is when there is an overall tilt in the wavefront as shown in figure 7. Since the true wavefront has no curvature, the expected curvature signal is zero. Although this is the case in the centre of the aperture, it is not the case at the edges. It is clear that the problem with this interpretation of the measurement data is that it fails to model the edge effects, and for this reason a practical curvature sensing instrument measures the wavefront slope separately.

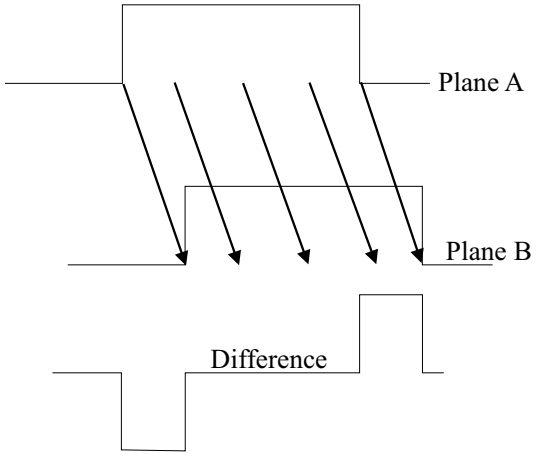

Figure 7: A wavefront that is only tilted has 0 curvature, but still produces an edge signal.

2.2 Estimating slopes

The measurement data can, however also be interpreted in terms of slopes [5]. The simplified 1D version of the propagation of light is shown in figure 8 where we have traced rays from points in the aperture using the assumption of geometric optics. Since the rays do not cross, if we integrate the light between the two rays P_1

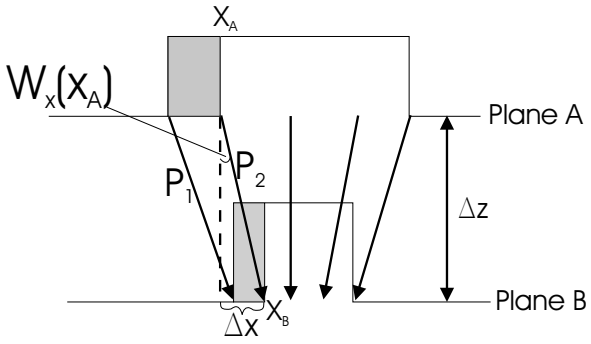

Figure 8: Derivation of the wavefront slope, using the same data from figure 3.

and P_2 at planes A and B (grayed region), the result must be equal.

This conservation of light during propagation may be expressed mathematically as,

$$C_{I_A}(x_A) = \int_{-\infty}^{x_A} I_A(x)dx = \int_{-\infty}^{x_B} I_B(x)dx = C_{I_B}(x_B) \qquad (6)$$

where $I_A(x)$ and $I_B(x)$ are the intensity distributions across planes A and B respectively.

The wavefront slope corresponding to ray P_2 is thus equal to $\frac{\Delta x}{\Delta z}$, where $\Delta x = x_B - x_A$ and Δz is the distance between the two planes.

Equation (6), is equivalent to histogram specification, a commonly used image processing technique and indeed if the intensity distributions are normalised to 1 they can be considered to be the probability distributions for photon arrival.

The wavefront slope across the entire aperture can thus be found by first converting the intensity distributions in the measurement planes to cumulative density functions, as shown in figure 9a. Histogram specification can then be used to estimate the wavefront slopes directly. The slopes can be differentiated to estimate the curvature or integrated to find the phase aberration as shown in figure 9a.

2.3 Comparison of the curvature and slope analyses

It is apparent from figure 9b that while geometric optics holds slope estimation produces an exact estimate of the curvature and overall phase for a pure defocus. This is in contrast to the results obtained from estimating the curvature directly. Similarly, when the phase aberration is a pure tilt, a direct estimation of the slope removes the edge effects that are apparent when estimating the curvature.

To see why this difference arises, it is worthwhile to consider the fundamental assumptions that underpin the two approaches. When estimating the curvature directly, it is implicitly assumed that the effect of the wavefront aberration is to move light from one measurement plane to another, whereas the slope estimation assumes that light is being displaced within a measurement plane.

In order to clarify this assumption implicit in the curvature sensor, consider the two stage process used to estimate slopes. The data is first differenced, and then integrated. Since these processes are linear, the order can be reversed to obtain figure 9b. Whereas slope estimation matches the two distributions along the abscissa, curvature estimation does it along the ordinate, resulting in erroneous results at the boundaries.

3 2D wavefront reconstruction

Although the slope based interpretation produces superior results in one dimension, there is some difficulty in generalising the problem to two dimensions. Because the light movement is not resticted to one-direction the radon transform must be employed to reduce the problem to a set of one-dimensional problems.

Figure 10 illustrates how the intensity data is first projected to form a one-dimensional intensity distribution. Histogram specification is then performed to estimate the line integral of the phase slopes taken along the same lines as the intensity data was projected. Provided the intensity data is projected at several angles it is possible to reconstruct the overall slope data using the same techniques that are employed in medical imaging[6].

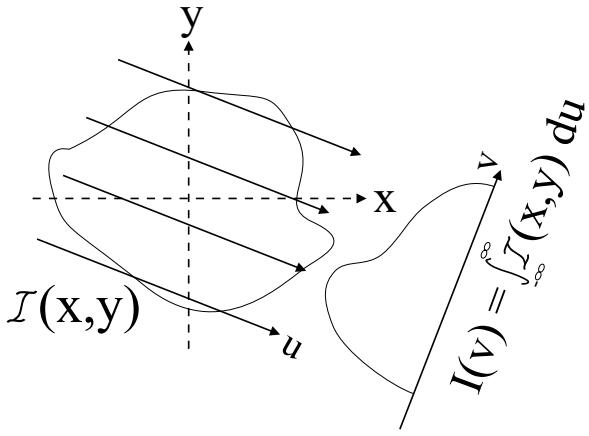

Figure 10: Projection of an intensity distribution onto the v-axis.

3.1 Basis function representation

A more convenient method for reconstruction relies on the linearity of the Radon transform. The phase aberrations are conveniently described using circular basis functions, allowing us to describe a wavefront as a weighted combination of known modes. The

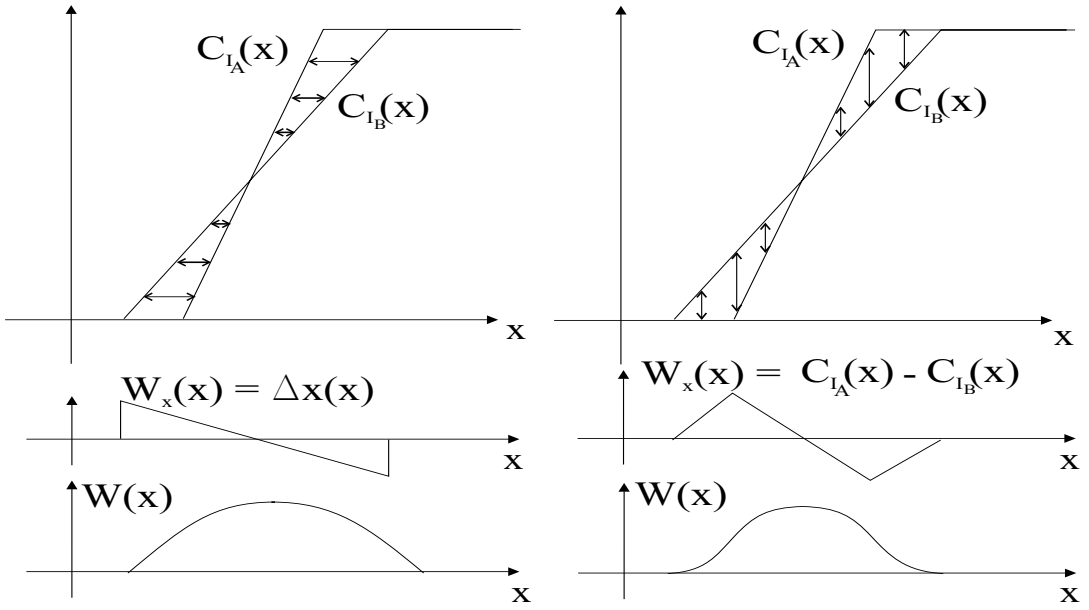

Figure 9: Comparison of the histogram specification process (a) with curvature sensing (b) in the estimation of slopes.

most common choice for the modes is the Zernike polynomials[7] since these relect the statistical eigenfunctions of atmospheric turbulence [8]. Hence,

$$W(x,y) = \sum_{i=1}^{\infty} \alpha_i Z_i(x,y) \tag{7}$$

where α_i represents the coefficient for each Zernike mode $Z_i(x,y)$.

Since differentiation is a linear process, if the phase aberration is a sum of Zernike polynomials, then the slopes estimated at each of the projections are proportional to the sum of derivatives of the Zernike polynomials. It is thus possible to estimate the coefficients of the Zernike polynomials using a system of linear equations, and solve for the coefficient of equation (7) using matrix techniques.

4 Demonstration of sensor

4.1 Experimental setup

Figure 11 shows the physical setup used for acquiring two out-of-focus images. A 633nm filtered and collimated laser beam is split into two identical paths. Two symmetrically out-of-focus intensity distributions are then measured at each path simultaneously.

The images obtained are shown in figure 12. The parameters for this setup are as shown in figure 11.

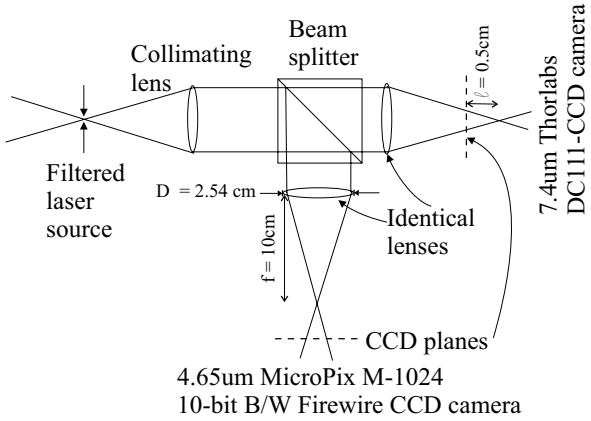

Figure 11: Experimental setup demonstrating the principles of wavefront sensing.

4.2 Recovery of wavefront

For a pair of intensity distributions, 8 different projections, with angles equally spaced between 0° and 180°, were histogram specified. The operations of differentiation (taking the wavefront slope), radon transform and vector stacking are linear, so a matrix (**H**) representing the forward problem may be calculated. The inverse of this process ($\mathbf{H}^{\dagger}$) may then be stored, and used to recover the wavefront during operation of the wavefront sensor. The first 14 Zernike modes were recovered. This should adequately model the expected aberrations.

In our setup, data from the experiment was processed offline in Matlab. After resampling the images to match pixel sizes, a data vector is formed from 8 projections and histogram specifications. The Zernike coefficients

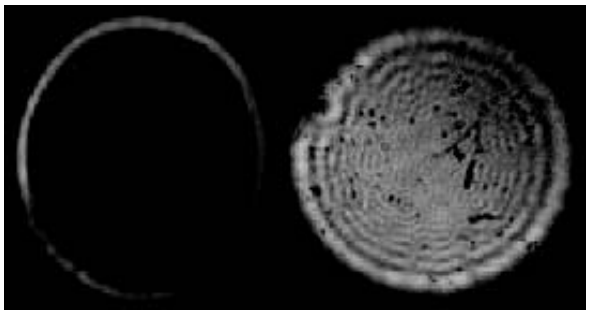

Figure 12: Intensity distributions measured in the post-focal plane (left), and pre-focal plane (right). The images have been resampled and normalised.

of the wavefront were then directly obtained from $\alpha = \mathbf{H}^{\dagger}\mathbf{d}$.

Figure 13 shows the coefficients of the recovered wavefront. The reconstructed wavefront, with tip and tilt terms removed for presentation purposes, is shown in figure 14. The terms Z_2, Z_3 and Z_4, representing the tip, tilt, and defocus terms respectively, are useful for calibrating an optical setup to reduce systematic wavefront errors. The higher order terms consisting of aberrations such as astigmatism and coma decay towards zero as expected.

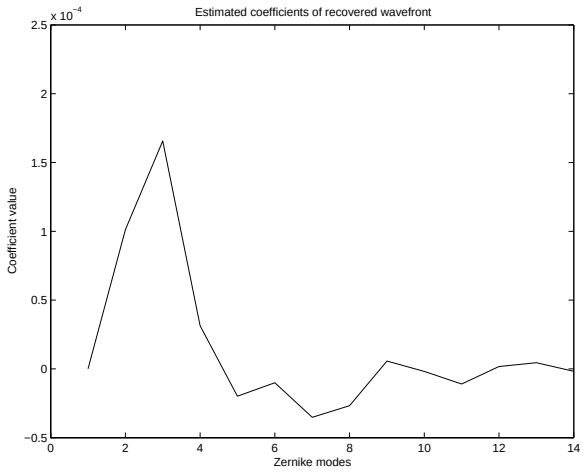

Figure 13: Estimated coefficients using the data from figure 12.

5 Conclusion

This paper outlined the wavefront reconstruction algorithms of the curvature and slope based wavefront sensor. We provided an intuitive explanation of the operation of the slope sensor, and showed how its unified treatment of intensity data is superior to that of the curvature.

An experimental demonstration of the slope based sensor with results is also presented to confirm its operation. The aberrations observed are due to errors in the optical alignment. This ability to reconstruct these errors provides a useful calibration tool.

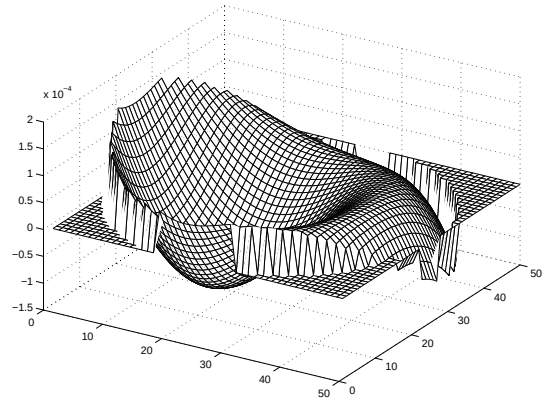

Figure 14: Wavefront reconstructed from the estimated coefficients of figure 13 with tip and tilt terms removed.

The slope based sensor successfully operates in open-loop, in contrast to the traditional curvature sensor, which requires a closed-loop operation under moderately severe optical aberrations. Although there is more processing required for the geometric wavefront sensor, this does not preclude real-time operation. Overall, the demonstration system has shown potential for implementation in a real-time adaptive optics system.

The authors acknowledge support from the Keith Laugesen Memorial scholarship, and the Marsden fund.

References

[1] F. Roddier. The effects of atmospheric turbulence in optical astronomy. *Progress in optics*, 19(5):281–376, 1981.

[2] M. A. van Dam and R. G. Lane. Tip/tilt estimation from defocused images. *Journal of the Optical Society of America A*, 19(4):745–752, Apr 2002.

[3] F. Roddier. Curvature sensing and compensation: a new concept in adaptive optics. *Applied Optics*, 27:1223–1225, 1988.

[4] M. A. van Dam and R. G. Lane. Extended analysis of curvature sensing. *Journal of the Optical Society of America A*, 19(7):1390–1397, July 2002.

[5] M. A. van Dam and R. G. Lane. Direct wavefront sensing using geometric optics. In *High Resolution Wavefront Control: Methods, Devices and Applications IV, Proceedings of SPIE*, volume 4825, 2002.

[6] Henry J. Scudder. Introduction to computer aided tomography. *Proceedings of the IEEE*, 66(6):628–637, June 1978.

[7] Robert J. Noll. Zernike polynomials and atmospheric turbulence. *Journal of the Optical Society of America*, 66(3):207–211, 1976.

[8] M. C. Roggemann and B. Welsh. *Imaging through turbulence*. The CRC Press, 1 edition, 1996.