

Optical misalignment sensing and image reconstruction using phase diversity

R. G. Paxman and J. R. Fienup

Optical Science Laboratory, Advanced Concepts Division, Environmental Research Institute of Michigan,
P.O. Box 8618, Ann Arbor, Michigan 48107

Received October 30, 1987; accepted January 11, 1988

A segmented-aperture telescope such as the Multiple-Mirror Telescope will suffer from phase errors unless the segments are aligned to within a small fraction of a wavelength. Such a coherent alignment of the segments is difficult to achieve in real time. An alternative is to record the images degraded by phase errors and to restore them after detection by using phase-retrieval techniques. In this paper we describe the use of Gonsalves's phase-diversity method (which was previously used to combat atmospheric turbulence) to correct imagery blurred by a misaligned segmented-aperture telescope. Two images are recorded simultaneously: the usual degraded image in the focal plane and a second degraded image in an out-of-focus plane. An iterative gradient-search algorithm finds the phase error of the telescope that is consistent with both degraded images. We refer to this technique as the method of multiple-plane measurements with iterative reconstruction. The final image is obtained by a Wiener-Helstrom filtering of the degraded image using the retrieved phase errors. The results of reconstruction experiments performed with simulated data including the effects of noise are shown for the case of random piston phase errors on each of six segments.

1. INTRODUCTION

The trend in telescope design in recent years has favored segmented-aperture systems. Examples of such systems that are either functioning or in various stages of development include the Multiple-Mirror Telescope (MMT),¹ the Keck telescope,² the Very Large Telescope,³ and the National New Technology Telescope.⁴ Multiple-segment telescopes have the potential to achieve resolutions far superior to that of a single segment while avoiding the considerable problems encountered in the fabrication of a monolithic primary mirror large enough to achieve a comparable resolution. In order to reach the full resolution potential, however, the individual segments must be aligned to within a small fraction of a wavelength. Thus, although mirror fabrication tolerances are relaxed by the segmented-aperture concept, the problem of alignment among segments is introduced. Considerable effort has been directed toward the development of active and passive methods for sensing piston (i.e., a different constant phase error over each segment) and tilt (i.e., a different linear phase error over each segment) misalignments. Active methods typically involve local laser interferometers that interfere light between adjacent segments.⁵ This approach requires considerable additional optical hardware that is itself subject to misalignment. One passive method relies on the presence of a bright star (with a magnitude of 8 or 9) near the object of interest.¹ Phasing between any two segments is performed by forming white-light fringes when imaging the bright star. Unfortunately, once the system is aligned, it must be redirected toward the object of interest, introducing flexure effects. Observation time is limited by alignment drift caused by thermal and gravitational effects.

In situ alignment sensing could be performed passively for extended objects by using a shearing interferometer, but,

again, a significant amount of additional hardware must be devoted solely to alignment sensing.⁶ The use of a Hartman sensor also has this drawback. In addition, the sensitivity of the Hartman sensor degrades as the objects being imaged become more extended.

In this paper we present an alternative to the methods surveyed above. The method described can be used for both aberration sensing and high-resolution image reconstruction. It relies on an external reference, the object being imaged, and should therefore be more robust than local interferometer alignment sensing or methods that rely on a nearby bright star. The technique also works well for extended objects, with no loss of sensitivity owing to object extent. Furthermore, the additional hardware required for the technique is modest and includes a simple beam splitter and a second detector array. Figure 1 depicts a cross section of the MMT, which serves as one example of a segmented-aperture system. Notice that a beam splitter has been introduced so that a second image can be collected. The second image is intentionally translated along the optical axis by a distance ΔZ so that a known amount of defocus is introduced. The goal is to infer the misalignment parameters from the two collected images, and this is performed with an iterative algorithm. The method is appropriately called the method of multiple-plane measurements with iterative reconstruction (MMIR). Once the misalignment parameters are known, they can be corrected actively. If active alignment correction is unavailable or if misalignment sensing cannot be performed fast enough to follow misalignment drift rates, a Wiener-Helstrom filter can be constructed to deblur the degraded imagery that is collected.

In Section 2 a mathematical model is described for the misalignment-sensing problem, and the image-reconstruction algorithm is presented. In Sections 3 and 4 reconstruc-

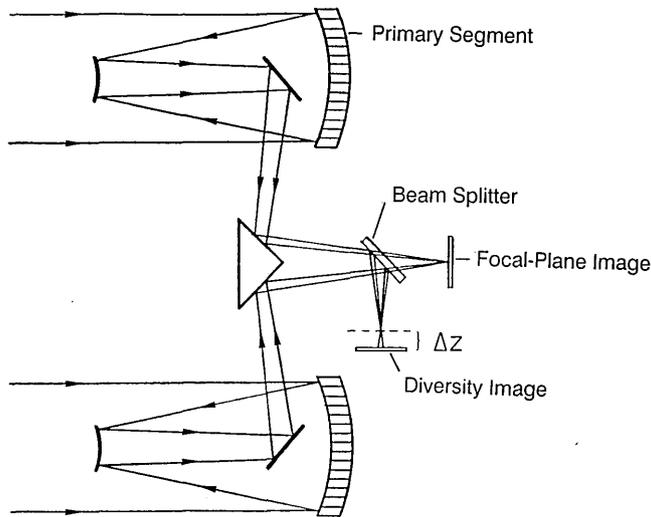


Fig. 1. Cross section of the MMT adapted for multiple-plane imaging.

tion results are shown for computer-simulated data for the cases of noise-free and noisy data, respectively. In Section 5 we present conclusions that can be drawn from the computer-simulation experiments.

2. FORMULATION OF THE PROBLEM AND THE RECONSTRUCTION ALGORITHM

We begin by making the simplifying assumption that the object or scene to be imaged is illuminated with spatially incoherent, quasi-monochromatic light. In addition, the imaging system is presumed to be a linear shift-invariant system, leading to the following imaging equation:

$$g(x) = f(x) * s(x), \quad (1)$$

where $g(x)$ is the given image, $f(x)$ is the object to be found, $s(x)$ is the point-spread function (PSF) of the system, x is a two-dimensional vector, and the asterisk represents a two-dimensional convolution. The PSF may have an unknown aberration associated with it that is due to system misalignment. The Fourier representation of Eq. (1) is given by

$$G(u) = F(u)S(u). \quad (2)$$

The optical transfer function (OTF) of the system, $S(u)$, is found by autocorrelation of the coherent transfer function (CTF),

$$S(u) = C(u) \star C(u), \quad (3)$$

where the star represents a cross correlation. For a segmented-aperture system the CTF takes the form

$$C(u) = \sum_{n=1}^N A_n(u) \exp[i\phi_n(u)], \quad (4)$$

where $A_n(u)$ is the binary aperture function representing the n th segment, $\phi_n(u)$ is the phase aberration associated with the n th segment, and N is the total number of segments. The method of MMIR requires that a second image be collected. This second image is intentionally defocused and therefore has a different CTF:

$$C_d(u) = \sum_{n=1}^N A_n(u) \exp\{i[\phi_n(u) + \Delta\phi(u)]\}, \quad (5)$$

where the subscript d indicates that a phase diversity has been applied to the system and $\Delta\phi(u)$ represents the known phase-diversity term. Because it corresponds to defocus, $\Delta\phi(u)$ will be quadratic in $|u|$, where $|u|$ denotes the length of the vector u . It is important to understand that although each $\phi_n(u)$ is unknown, $\Delta\phi(u)$ is known because the exact amount of defocus can be determined by the location of the second image plane relative to the first. The diversity OTF is the autocorrelation of the diversity CTF:

$$S_d(u) = C_d(u) \star C_d(u). \quad (6)$$

Now consider the Fourier representation of the imaging equations for the two collected images together:

$$G(u) = S(u)F(u), \quad (7)$$

$$G_d(u) = S_d(u)F(u). \quad (8)$$

Equations (7) and (8) illustrate the problem that we wish to solve. Both $G(u)$ and $G_d(u)$ are given by the image data. $S(u)$, $S_d(u)$, and $F(u)$ are unknowns that we wish to find. Fortunately, we have additional knowledge about the relationship between $S(u)$ and $S_d(u)$ through our knowledge of the phase diversity. Suppose now that we make an estimate of the object's Fourier transform, $\hat{F}(u)$, and of the two OTF's, $\hat{S}(u)$ and $\hat{S}_d(u)$. The mean-squared difference between the data predicted by our estimates and that actually collected serves as a measure of the quality of our estimates:

$$E = \sum_u |G(u) - \hat{S}(u)\hat{F}(u)|^2 + \sum_u |G_d(u) - \hat{S}_d(u)\hat{F}(u)|^2. \quad (9)$$

For given $\hat{S}(u)$ and $\hat{S}_d(u)$, it can be shown that this error metric may be minimized by the choice of an object estimate whose Fourier transform is given by

$$\hat{F}(u) = \frac{\hat{S}^*(u)G(u) + \hat{S}_d^*(u)G_d(u)}{|\hat{S}(u)|^2 + |\hat{S}_d(u)|^2}. \quad (10)$$

Substitution of Eq. (10) back into Eq. (9) with appropriate algebraic manipulations yields

$$E = \sum_u \frac{|G(u)\hat{S}_d(u) - G_d(u)\hat{S}(u)|^2}{|\hat{S}(u)|^2 + |\hat{S}_d(u)|^2}. \quad (11)$$

Remarkably, Eq. (11) expresses a figure of merit for the correctness of the OTF estimates that is independent of any object estimate. This error metric was first derived by Goncalves,^{7,8} who used it to correct phase aberrations caused by atmospheric turbulence for single-aperture systems.

Misalignment sensing can be accomplished as follows. Because the phase aberrations may be parameterized (e.g., with piston and tilt parameters), a given parameter-set estimate allows for the computation of the corresponding estimates of the conventional and diversity CTF's and subsequently the OTF's by means of Eqs. (3)–(6). Consequently, the error metric in Eq. (11) can be computed. This error metric is therefore defined on a multidimensional parameter space, the coordinates of that space being the unknown misalignment parameters.

Standard nonlinear optimization techniques can be used to find the coordinates of the metric's global minimum, corresponding to the actual phase-aberration parameters. An initial estimate for the parameter set is made. In the absence of additional information the aligned configuration (i.e., the configuration with no phase errors) is a natural choice. The goal is to find the parameter set for which the error metric, or objective function [Eq. (11)], is a minimum. We can evaluate the objective function and the gradient of the objective function at the initial estimate. In our simulations the gradient was approximated by the method of finite differences, but an analytic expression for the gradient that is computationally more efficient could also be used, as described in Appendix A. The gradient provides guidance for choosing a direction to move in parameter space that reduces the objective-function value. We computed this direction by using the conjugate-gradient method, which is known to have far better convergence rates than the method of steepest descent for quadratic objective functions.⁹ Once the direction is established, a line search is performed to find the minimum objective-function value along the search direction. This is done by evaluating the objective function at several locations along the line-search direction and performing a cubic interpolation to find the minimum. This new location in parameter space takes on the role of the current estimate in parameter space, and the entire procedure is repeated iteratively. When the change in consecutive current estimates of the parameters is very small and the norm of the gradient is small as well, the iterations are terminated, since the algorithm has probably found a local minimum.

The optimization procedure produces parameter estimates that can be used to drive an adaptive-optics compensation system. Alternatively, the knowledge of the phase aberrations can be used to reconstruct a high-resolution image from the raw data. We employ a Wiener-Helstrom filter¹⁰ to perform the object restoration:

$$\hat{F}(u) = \frac{G(u)S^*(u)}{|S(u)|^2 + \frac{P_n(u)}{P_f(u)}}, \tag{12}$$

where the second term in the denominator is the ratio of the power spectral density of the noise to that of the object class. The reconstructed image is then given by the inverse Fourier transform of $\hat{F}(u)$. In practice the power spectral density of the noise is frequently assumed to be a constant. The power spectral density of the object class is usually unavailable and must be estimated. For the results shown in Sections 3 and 4 we simply used an appropriate constant for the ratio of power spectra.

3. SIMULATIONS

Computer simulations were used to test the method of MMIR. A six-segment MMT-like aperture was designed to fit within a 64×64 complex array, as shown in Fig. 2. Various amounts of piston phase were added to each of the six segments. A random set of misalignment parameters of particular interest is shown in Fig. 3. This particular misalignment configuration, consisting of only piston terms, was used as the actual configuration, or true solution, in all

the simulations presented here. The corresponding OTF was created by autocorrelating the CTF, which was accomplished by embedding the CTF in a 128×128 complex array, performing a fast Fourier transform (FFT), taking the modulus squared, and applying an inverse FFT.

The diversity CTF was formed by adding a global quadratic phase to the conventional CTF, and the diversity OTF was formed by autocorrelating the diversity CTF. In practice the quadratic phase diversity depends on the wavelength, λ , the focal length of the system, Z_f , and the distance that the diversity image is translated along the optical axis, ΔZ :

$$\Delta\phi(u) = \frac{\pi\Delta Z}{\lambda Z_f^2} |u|^2, \tag{13}$$

where $|u|$ is the length of the vector u . For Eq. (13) it is assumed that the origin of the coordinate system defining the vector u coincides with the optical axis, that is, the origin is located at the center of the six-segment array. The number of waves of quadratic diversity at the edge of the entire six-segment aperture is given by

$$\text{Number of waves of defocus} = \frac{\Delta Z}{8\lambda(F\#)^2}, \tag{14}$$

where $F\#$ represents the F number (focal length/diameter ratio) of the system. In the experiments reported here we used 0.5 and 1.0 waves of diversity at the edge of the aperture. Modulation transfer functions, i.e., the moduli of the corresponding OTF's, for aligned and misaligned systems with and without diversity are shown in Fig. 4.

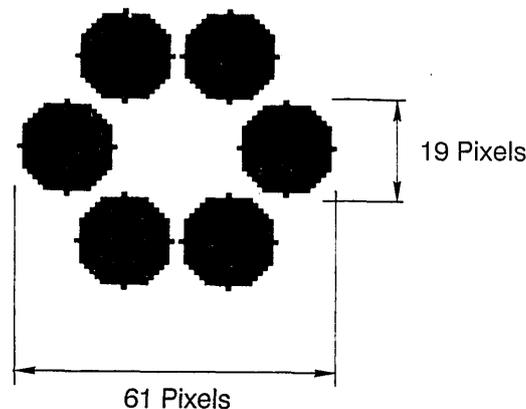


Fig. 2. Simulation of the aperture for a MMT-like system.

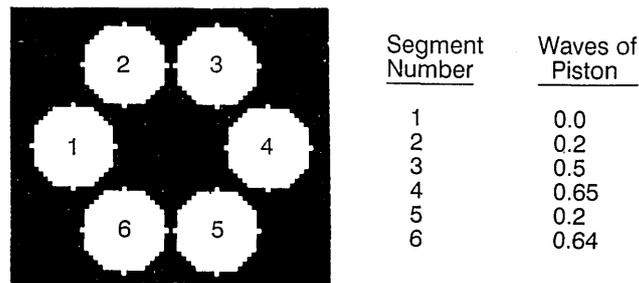


Fig. 3. Piston configuration representing a true solution for simulations.

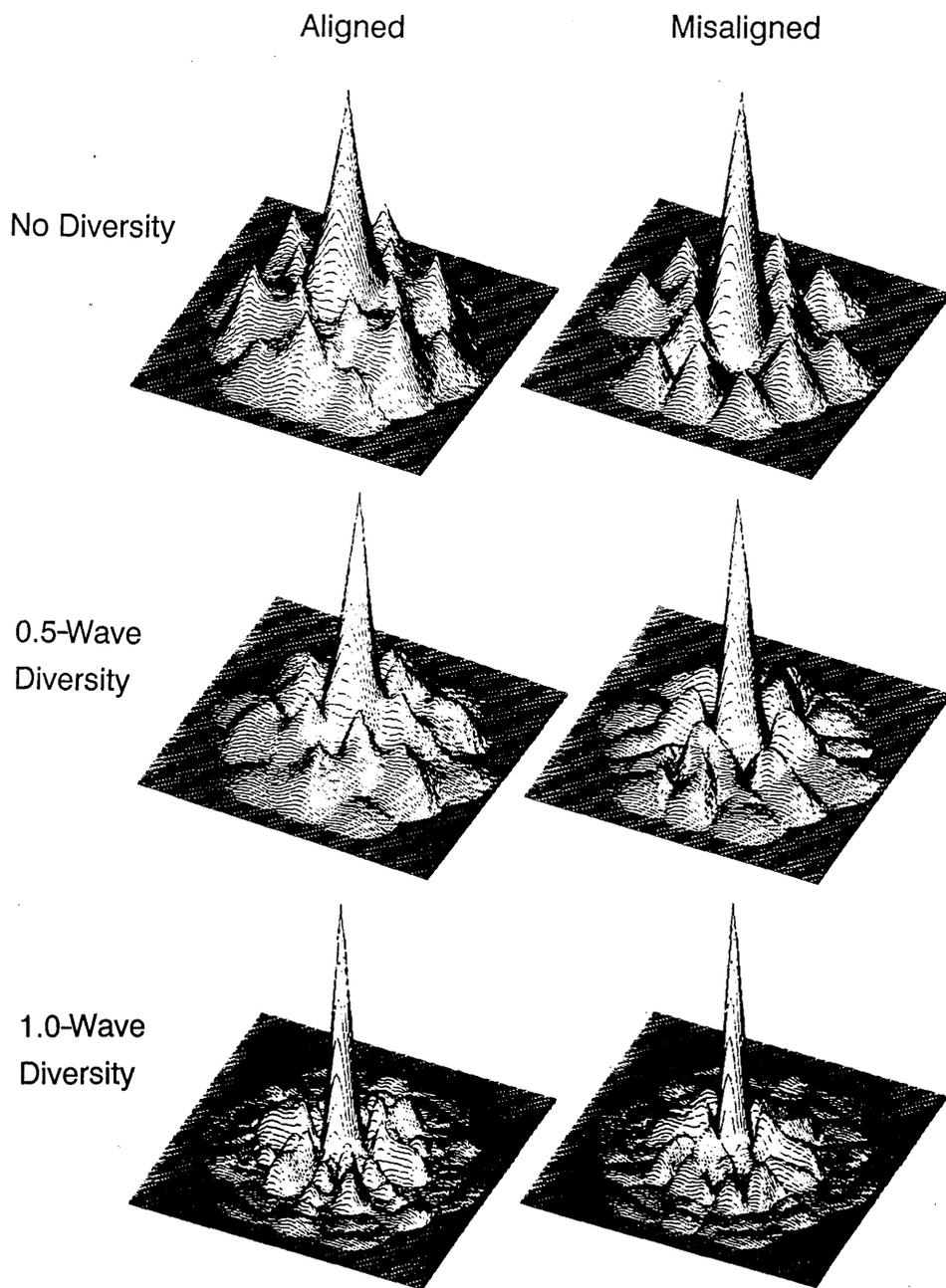


Fig. 4. Modulation transfer functions for aligned and misaligned systems with and without phase diversity.

A 128×128 Voyager image of Jupiter was used to represent the true object and is shown in Fig. 5(A). First, to see what would happen if the system were aberration free, we simulated two data images by imaging with the conventional OTF and its diversity counterpart for the case of an aligned system. These images are shown in Figs. 5(B) and 5(C). The amount of diversity used here was 0.5 wave. Figure 5(D) shows the result of applying a Wiener-Helstrom filter to the ideal image shown in Fig. 5(B). The filtering is effective in correcting the OTF attenuation that is due to the sparse aperture. Next, we simulated the data for the aberrated system. Figures 5(E) and 5(F) correspond to the conventional and diversity images collected with the misaligned system. With these two data images in place, the optimization procedure can begin. When the procedure described in

Section 2 was applied to the two data images associated with the misaligned system, using parameters for an aligned system as an initial estimate, the restored misalignment parameters were virtually perfect (each within 10^{-5} wave). These restored parameters were used to construct a Wiener-Helstrom filter that was used to deblur the simulated data image shown in Fig. 5(E). The resulting restoration, shown in Fig. 5(H), is visually indistinguishable from the true object. It is interesting to construct a Wiener-Helstrom filter for an aligned system (assuming zero phase error) and apply it to the images collected with the misaligned system. We call this the naive restoration because, in the absence of knowledge of the misalignment parameters, one might naively apply this filter. The naive restoration is shown in Fig. 5(G). Clearly the image shown in Fig. 5(H) is superior to

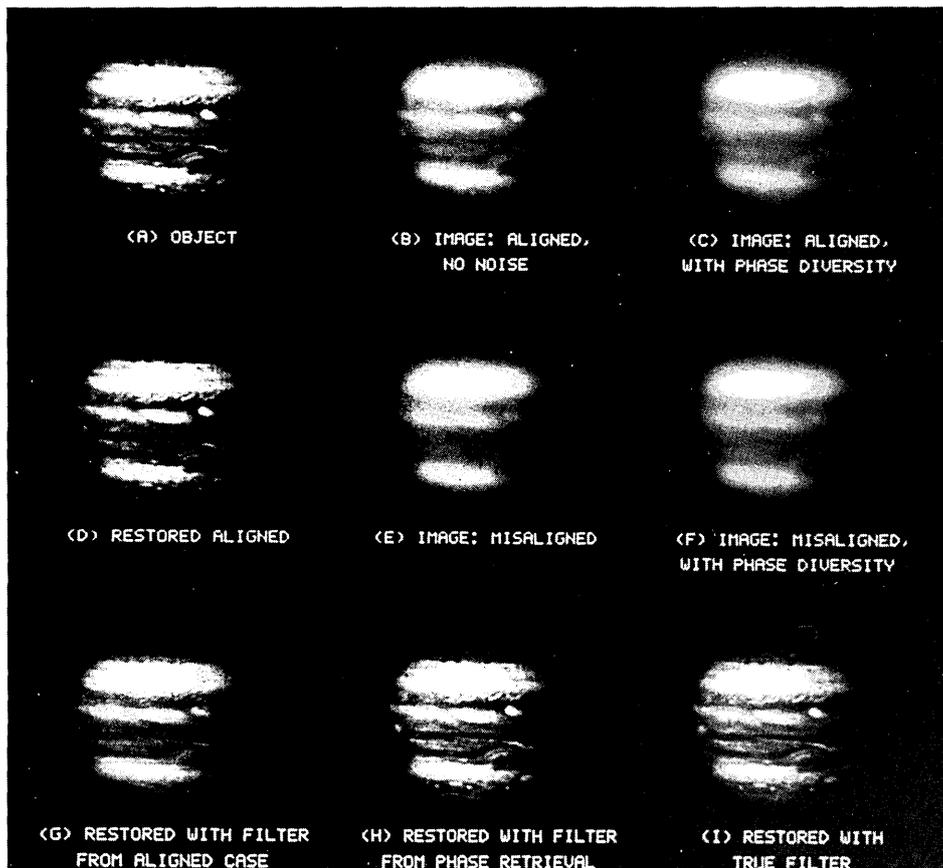


Fig. 5. Noiseless-data images and reconstructions: (A) original object (Jupiter), (B) image of object through aligned system, (C) diversity image for aligned system (0.5-wave diversity), (D) Wiener-Helstrom restoration of (B), (E) image through misaligned system, (F) diversity image for misaligned system, (G) restoration of (E) obtained by using a Wiener-Helstrom filter for an aligned system, (H) restoration of (E) obtained by using a Wiener-Helstrom filter constructed with MMIR parameter estimates, and (I) restoration of (E) obtained by using a Wiener-Helstrom filter constructed with true misalignment parameters.

those shown in Figs. 5(G) and 5(E), thus demonstrating the utility of MMIR. Finally, Fig. 5(I) shows a restoration of the image shown in Fig. 5(E) obtained by using the true misalignment parameters. The result is indistinguishable from our restoration in Fig. 5(H), since the restored misalignment parameters are so close to the true misalignment parameters.

One of the concerns with using a gradient-search algorithm is the possibility of becoming entrapped in a local minimum that is not the true solution. One way to avoid this problem is to perform the minimization many times with different initial estimates each time. The hope is that at least one of the minimizations will lead to the true solution. In order to explore the need for such a strategy, the entire parameter space was coarsely sampled. Recall that there are six segments to which a piston aberration can be added. Of these, however, only five are independent, since a constant phase error over all segments will have no effect on the imagery or the objective function. The parameter space therefore has five dimensions. In addition, the assumption of quasi-monochromaticity implies that each of the dimensions is periodic with a period of 2π rad or 1.0 wave.

Each parameter was sampled with a sampling interval of $\pi/2$ rad, giving four samples per parameter for a total of $4^5 = 1024$ samples. The objective function was evaluated at each

of the samples. These samples were then rank ordered on the basis of their objective-function values, and several of the sample points were used subsequently as initial estimates in a minimization sequence. Of the nine initial estimates, all but one led to the true solution. The details of the ranking of the initial estimates and their Euclidian distances from the true solution in parameter space are given in Table 1. It can be shown that, for this six-segment aperture and for only piston errors, the rms phase difference between two

Table 1. Multiple-Initial-Estimate Experiment Results

Rank Order for Initial Estimate	Objective Function ($\times 10^6$)	Euclidian Distance from True Solution	Type of Minimum Found
1	0.17119	0.16462	Global
2	0.21210	0.29086	Global
3	0.32122	0.37363	Global
4	0.36984	0.56754	Global
5	0.53035	0.44956	Global
10	0.65025	0.79662	Global
50	1.0932	0.27767	Global
513	2.1779	0.66866	Global
1024	3.8361	0.62016	Local

configurations is proportional to the Euclidian distance in parameter space with a proportionality constant of $\sqrt{6}/6$. The results in Table 1 suggest that the shape of the objective function for this particular misalignment is relatively well suited to gradient-search methods. A diversity of 1.0 wave was used for this experiment.

4. SIMULATIONS WITH NOISE

It is difficult to visualize and characterize the morphology of a multidimensional objective function that has many alignment parameters as the underlying independent variables. However, it is possible to plot the error metric as a function of a one-dimensional cut through parameter space in a specified direction. We use this probe to check how the surface contour of the objective function changes when the data are corrupted with additive noise.

Figure 6(A) gives a one-dimensional plot of the objective function when a single piston parameter is varied from the true solution. The objective function was computed by using conventional and diversity images with 1.0% additive zero-mean Gaussian white noise, where noise is specified by the ratio of the standard deviation of the noise to the peak value in the noiseless focal-plane image. Whereas the gross surface contour illustrated in Fig. 6(A) is consistent with a well-behaved objective function, the detailed structure is troublesome. The high-frequency structure suggests the existence of many local minima far from the global minimum, and it provokes the question of how close to the true solution an initial estimate must be for a minimization sequence to find the global minimum. Figures 6(B) and 6(C) show magnified versions of the same one-dimensional cut of the objective function, with magnifications of 10 \times and 100 \times ,

respectively. Surprisingly there is no smoothing of the objective function at these magnifications. In other words, the objective function appears to have a fractal-like structure in the presence of noise. Care must be used in interpreting these one-dimensional cuts. A local minimum in a one-dimensional cut need not imply a local minimum in the full multidimensional objective function. Still, the fractal-like structure is strong evidence for the existence of many local minima. In practice it was observed that the minimizer stagnated away from the true solution when 1.0% noise was added to the data. This is consistent with the conjecture that noise creates local minima in the objective function.

Clearly the presence of noise in the data presents a problem for the MMIR procedure. We present here a simple technique that was developed to suppress these troublesome noise effects. Recall that the error metric is computed according to Eq. (11) and that, of the quantities that appear in that expression, only G and G_d are affected by noise. Furthermore, since the collected images have passed through the true OTF's of the system before the addition of noise, the signal-to-noise ratios at the high spatial frequencies of G and G_d will be considerably less than those at the lower spatial frequencies. This observation suggests a simple truncation of the summation in Eq. (11) so that the limits of the summation span only lower spatial frequencies, thus reducing the effects of noise. A particular truncation that we used was equivalent to performing the summation over a 24×24 pixel subframe centered on the dc value. A one-dimensional cut for the truncated error metric applied to the same noisy data is shown in Fig. 6(D). The improvement in the shape of the objective function is dramatic and encouraging. Notice that the minimum along this one-dimensional cut has shifted by approximately one thousandth of a wave. Of course it is

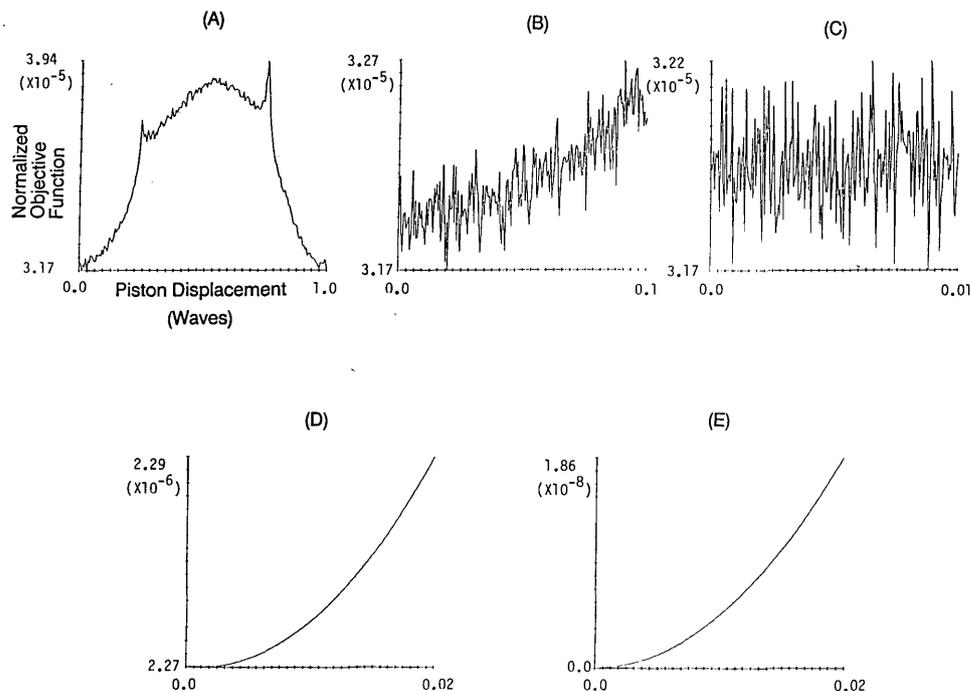


Fig. 6. One-dimensional cuts of normalized objective function created by varying a single piston parameter: (A) variation over one wave with noisy data, (B) structure at a magnification of 10 \times , (C) structure at a magnification of 100 \times , (D) objective function resulting from truncated error metric with noisy data, and (E) objective function resulting from truncated error metric with noiseless data.

Table 2. Reconstruction Parameters for 1% Noise^a

Segment Number	Final Estimate (wave)	True Solution (wave)
1	0.0	0.0
2	0.188	0.2
3	0.486	0.5
4	0.635	0.65
5	0.208	0.2
6	0.632	0.64

^a The initial estimate was zero for each segment.

difficult to infer from this how far the minimum of the multidimensional objective function has shifted from the true solution. That the shift is due to noise and not to the truncation operation alone is supported by Fig. 6(E), which shows the analogous one-dimensional cut for the truncated error metric applied to noiseless data. Notice that the minimum has not shifted from the true solution in this case.

It is interesting to note that the operation of truncation is equivalent to low-pass filtering of both data images before computation of an error metric. Apparently the operation

suppresses high-frequency structure imposed on the objective function by noise while still providing discrimination between imperfect OTF estimates and the true OTF's. Furthermore, the procedure actually reduces the number of operations required to compute an error metric. Interestingly enough, when the technique was used with noiseless data and the same initial estimate that led to a local-minimum solution in the multiple-initial-estimate experiment (Table 1), the global minimum was actually achieved. In practice, it was found necessary to use the technique to achieve successful minimizations even in the presence of small amounts of noise. Table 2 shows the results of a minimization for which 1% Gaussian noise was added to both data images. The restored parameters are each within 0.015 wave of the true solution, and the rms phase error over the total aperture is 0.0094 wave.

A series of collected images and reconstructions similar to those of Fig. 5, but for the 1% noise case, is given in Fig. 7. As before, the MMIR restoration for the misaligned system [Fig. 7(H)] compares favorably with the corresponding data image [Fig. 7(E)] and the naive image [Fig. 7(G)]. These images demonstrate that MMIR can be used successfully in

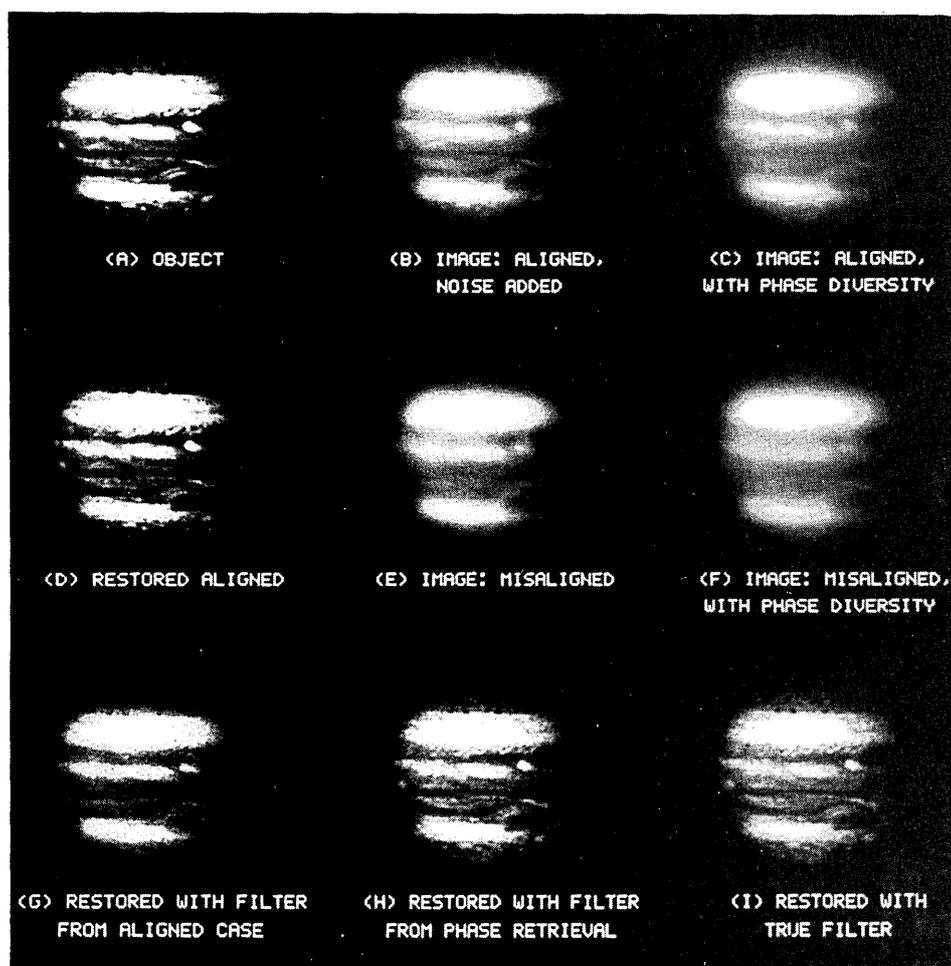


Fig. 7. Data images with 1% noise added and reconstructions: (A) original object, (B) image of object through aligned system, (C) diversity image for aligned system (0.5-wave diversity), (D) Wiener-Helstrom restoration of (B), (E) image through misaligned system, (F) diversity image for misaligned system, (G) restoration of (E) obtained by using a Wiener-Helstrom filter for an aligned system, (H) restoration of (E) obtained by using a Wiener-Helstrom filter constructed with MMIR parameter estimates, (I) restoration of (E) obtained by using a Wiener-Helstrom filter constructed with true misalignment parameters.

the presence of noise. It must be noted, however, that the image shown in Fig. 7(H) is not so true to the image shown in Fig. 7(D), as was the case for the corresponding noiseless data. In addition, Fig. 7(H) is comparable with Fig. 7(I), which represents the restoration of the noisy data obtained by using a filter constructed with absolute knowledge of the alignment parameters. Noise apparently presents a limitation on the fidelity of restored imagery even when the alignment parameters are known perfectly. This is because a loss in the signal-to-noise ratio is always incurred by a misaligned incoherent system.¹¹ This becomes evident when the aligned and misaligned MTF's are compared (Fig. 4). Clearly, the misaligned MTF attenuates the higher spatial frequencies more severely than does the aligned MTF. Problems associated with Wiener-Helstrom filtering of the data would be reduced considerably if MMIR were used in an adaptive mode to correct for misalignments.

5. CONCLUSIONS

The technique of using phase diversity to infer phase aberrations from two incoherent images, first developed by Goncalves,^{7,8} is demonstrated to be effective for determining segmented-aperture misalignment parameters. We refer to this method as multiple-plane measurements with iterative reconstruction (MMIR). Also demonstrated is image reconstruction by the use of the determined parameters. Segmented-aperture misalignment is easily parameterized with relatively few parameters, which helps to keep the computations required for a gradient search in parameter space manageable. An analytic expression for the gradient of the objective function is derived and promises significant computational savings in special cases. Computer simulations are

where the dependence on u has been dropped for simplicity. The functions G and G_d are simply the discrete Fourier transforms of the two data images. These Fourier transforms need be computed only once, and G and G_d may be saved for repeated use. In contrast, OTF estimates \hat{S} and \hat{S}_d must be recomputed for each new location in parameter space. In order to compute a single gradient by the method of finite differences, a total of $4(P+1)$ forward or inverse FFT's of size $L \times L$ are performed, where P is the number of independent parameters. For the results shown here, $P=5$ and $L=128$. When \hat{S} , \hat{S}_d , G , and G_d are combined to form the objective-function value, some savings in storage and computation can be realized by using the fact that these quantities are all Hermitian and that their squared moduli are centrosymmetric functions. Therefore only half of the $L \times L$ complex elements for each of these quantities need to be stored or combined.

An alternative to calculating the discrete gradient, with its repeated calculations of the objective function, is to employ an analytic expression for the continuous gradient. A single term in the gradient is given by

$$\frac{\partial E}{\partial k_{mn}} = \sum_u \frac{\partial}{\partial k_{mn}} \frac{|G\hat{S}_d - G_d\hat{S}|^2}{|\hat{S}|^2 + |\hat{S}_d|^2} \quad (\text{A2})$$

$$= \sum_u \frac{\partial}{\partial k_{mn}} \frac{(G\hat{S}_d - G_d\hat{S})(G^*\hat{S}_d^* - G_d^*\hat{S}^*)}{\hat{S}\hat{S}^* + \hat{S}_d\hat{S}_d^*}, \quad (\text{A3})$$

where k_{mn} is the m th misalignment parameter of the n th segment. If we perform the partial derivative in Eq. (A3) and then multiply out the resulting factors, we get an expression consisting of 32 terms. These terms collapse into the following concise expression:

$$\frac{\partial E}{\partial k_{mn}} = \sum_u \frac{\left(\hat{S}_d \frac{\partial \hat{S}}{\partial k_{mn}} - \hat{S} \frac{\partial \hat{S}_d}{\partial k_{mn}} \right) (G\hat{S}^* + G_d\hat{S}_d^*) (G_d^*\hat{S}^* - G^*\hat{S}_d^*) + \text{c.c.}}{(|\hat{S}|^2 + |\hat{S}_d|^2)^2}, \quad (\text{A4})$$

presented that demonstrate the MMIR technique successfully on a six-segment system for which only piston misalignments are allowed. Results of experiments with multiple initial estimates suggest that the surface contour of the objective function is appropriate for gradient-search methods. Whereas noise in the data introduces a fractal-like structure into the objective function, low-pass filtering the noisy images before computing the objective function smooths this structure. Simulations are presented that indicate that this prefiltering procedure makes it possible to use the MMIR technique successfully when noise is present.

APPENDIX A: COMPUTATION OF THE GRADIENT

The numeric computation involved in the gradient-search algorithm is considerable. Recall that the merit function is evaluated by performing the following computation:

$$E = \sum_u \frac{|G\hat{S}_d - G_d\hat{S}|^2}{|\hat{S}|^2 + |\hat{S}_d|^2}, \quad (\text{A1})$$

where c.c. denotes a complex conjugate. Notice that Eq. (A4) is a general expression. It holds for any aberration that can be parameterized. In addition, the dependence on the parameter is confined to the first factor. Therefore only the first factor needs to be computed for each parameter in the determination of the gradient. The remaining expression need be found only once per gradient computation.

We consider now the specific case of an aperture consisting of N circular segments and evaluate the partial derivative of the OTF with respect to a parameter. The CTF for the N -segment system can be expressed as follows:

$$C(u) = \sum_{n=1}^N \text{Circ}(u - u_n) \exp \left[i \sum_{m=1}^M k_{mn} \theta_m(u - u_n) \right], \quad (\text{A5})$$

where $\text{Circ}(u)$ is the pupil function of a single circular segment with radius r ,

$$\text{Circ}(u) = \begin{cases} 1 & \text{if } |u| \leq r; \\ 0 & \text{otherwise;} \end{cases} \quad (\text{A6})$$

$\theta_m(u)$ is the m th of M orthonormal basis functions that span the class of aberrations under consideration over a single

segment; and k_{mn} is the dimensionless coefficient of the m th orthonormal basis function for the n th segment. As an example, if the segments can be considered to be truly rigid bodies, then these basis functions will span piston and tilt phase aberrations. In this case there will be one piston and two tilt basis functions so that $M = 3$.

The corresponding OTF is, of course,

$$S = C \star C. \tag{A7}$$

Substituting Eq. (A5) into Eq. (A7) yields

$$S(u) = \int \left\{ \sum_{n=1}^N \text{Circ}(u' - u_n) \exp \left[i \sum_{m=1}^M k_{mn} \theta_m (u' - u_n) \right] \right\} \times \left\{ \sum_{n'=1}^N \text{Circ}(u' - u_{n'} - u) \right. \\ \left. \times \exp \left[-i \sum_{m'=1}^M k_{m'n'} \theta_{m'} (u' - u_{n'} - u) \right] \right\} du'. \tag{A8}$$

Taking the partial derivative of Eq. (A8) with respect to the specific parameter k_{pq} gives

$$\frac{\partial S}{\partial k_{pq}} = \int \text{Circ}(u' - u_q) i \theta_p (u' - u_q) \\ \times \exp \left[i \sum_{m=1}^M k_{mq} \theta_m (u' - u_q) \right] \\ \times \left\{ \sum_{n'=1}^N \text{Circ}(u' - u_{n'} - u) \right. \\ \left. \times \exp \left[-i \sum_{m'=1}^M k_{m'n'} \theta_{m'} (u' - u_{n'} - u) \right] \right\} du' \\ - \int \text{Circ}(u' - u_q - u) i \theta_p (u' - u_q - u) \\ \times \exp \left[-i \sum_{m'=1}^M k_{m'q} \theta_{m'} (u' - u_q - u) \right] \\ \times \left\{ \sum_{n=1}^N \text{Circ}(u' - u_n) \right. \\ \left. \times \exp \left[i \sum_{m=1}^M k_{mn} \theta_m (u' - u_n) \right] \right\} du'. \tag{A9}$$

It is useful to define the following function:

$$B_{pq}(u) \equiv \text{Circ}(u - u_q) i \theta_p (u - u_q) \exp \left[i \sum_{m=1}^M k_{mq} \theta_m (u - u_q) \right]. \tag{A10}$$

Equation (A9) may now be written in a more succinct form:

$$\frac{\partial S}{\partial k_{pq}} = (B_{pq} \star C) + (C \star B_{pq}). \tag{A11}$$

Cross correlation is not a commutative operation. Instead, it is easy to show that

$$(C \star B_{pq})(u) = [(B_{pq} \star C)(-u)]^*. \tag{A12}$$

Thus only one cross correlation needs to be performed in order to compute Eq. (A11).

In the special case for which misalignments are confined to piston errors, the function B_{pq} becomes $B_q = i \theta \text{Circ}(u - u_q) \exp(i k_q \theta)$, where θ is a constant equal to the value of the piston basis function for a single segment. The function B_q has the interpretation of being proportional to the CTF of the q th segment. In addition, it is a simple matter to show that

$$B_q \star C = i \theta \sum_{n=1}^N \exp[i \theta (k_q - k_n)] \\ \times \text{Circ}(u - u_q) \star \text{Circ}(u - u_n) \\ = i \theta \sum_{n=1}^N \exp[i \theta (k_q - k_n)] \\ \times (\text{Circ} \star \text{Circ})(u + u_n - u_q). \tag{A13}$$

Equation (A13) implies that the autocorrelation of the circle function, $\text{Circ}(u)$, could be calculated only once and then stored. The quantity $B_q \star C$ can therefore be found by adding versions of this archived function that are shifted and multiplied by constants. This eliminates the need for FFT's in the calculation of the partial derivative

$$\frac{\partial S}{\partial k_q}.$$

In this special case of only piston errors, the total forward or inverse FFT count in finding the continuous gradient is reduced to 4, as opposed to the $4N$ count required to estimate the gradient by the method of finite differences. As before, careful consideration of the quantities in Eq. (A4) that are either Hermitian or anti-Hermitian permits a reduction in required storage and operations.

Although the continuous-gradient approach has not yet been implemented, it promises considerable computational savings over the discrete-gradient approach in the case for which only piston misalignments are allowed.

REFERENCES

1. E. K. Hege, J. M. Beckers, P. A. Strittmatter, and D. W. McCarthy, "Multiple Mirror Telescope as a phased array telescope," *Appl. Opt.* **24**, 2565-2576 (1985).
2. T. S. Mast and J. E. Nelson, "Status report on the W. M. Keck Observatory and Ten Meter Telescope," in *Advanced Technology Optical Telescopes III*, L. D. Barr, ed., Proc. Soc. Photo-Opt. Instrum. Eng. **628**, 204-206 (1986).
3. D. Enard, "The ESO Very Large Telescope project: present status," in *Advanced Technology Optical Telescopes III*, L. D. Barr, ed., Proc. Soc. Photo-Opt. Instrum. Eng. **628**, 221-226 (1986).
4. K. L. Shu and S. Eisenberg, "Planning the National New Technology Telescope (NNTT): I. Optical designs," in *Advanced*

- Technology Optical Telescopes III*, L. D. Barr, ed., Proc. Soc. Photo-Opt. Instrum. Eng. 628, 66-79 (1986).
5. R. R. Butts, S. J. Cusumano, J. S. Fender, and C. R. DeHainaut, "Phasing concept for an array of mutually coherent laser transmitters," *Opt. Eng.* 26, 553-558 (1987).
 6. J. C. Wyant, "Use of an ac heterodyne lateral shear interferometer with real-time wavefront correction systems," *Appl. Opt.* 14, 2622-2626 (1975).
 7. R. A. Gonsalves and R. Childlaw, "Wavefront sensing by phase retrieval," in *Applications of Digital Image Processing III*, A. G. Tescher, ed., Proc. Soc. Photo-Opt. Instrum. Eng. 207, 32-39 (1979).
 8. R. A. Gonsalves, "Phase retrieval and diversity in adaptive optics," *Opt. Eng.* 21, 829-832 (1982).
 9. D. G. Luenberger, *Linear and Nonlinear Programming* (Addison-Wesley, Reading, Mass., 1984).
 10. C. W. Helstrom, "Image restoration by the method of least squares," *J. Opt. Soc. Am.* 57, 297-303 (1967).
 11. J. W. Goodman, *Introduction to Fourier Optics* (McGraw-Hill, San Francisco, 1968).