Extended capture range for focus-diverse phase retrieval in segmented aperture systems using geometrical optics

Alden S. Jurling and James R. Fienup*

The Institute of Optics, University of Rochester, Rochester, New York 14627, USA
**Corresponding author: fienup@optics.rochester.edu*

Received November 25, 2013; revised January 28, 2014; accepted January 29, 2014;
posted January 29, 2014 (Doc. ID 201616); published February 27, 2014

Extending previous work by Thurman on wavefront sensing for segmented-aperture systems, we developed an algorithm for estimating segment tips and tilts from multiple point spread functions in different defocused planes. We also developed methods for overcoming two common modes for stagnation in nonlinear optimization-based phase retrieval algorithms for segmented systems. We showed that when used together, these methods largely solve the capture range problem in focus-diverse phase retrieval for segmented systems with large tips and tilts. Monte Carlo simulations produced a rate of success better than 98% for the combined approach.

OCIS codes: (010.7350) Wave-front sensing; (110.6770) Telescopes; (100.5070) Phase retrieval; (080.0080) Geometric optics.
http://dx.doi.org/10.1364/JOSAA.31.000661

1. INTRODUCTION

Wavefront sensing by focus-diverse phase retrieval employs defocused images of point-like stars as measured point spread functions (PSFs) for determining the aberrations of a telescope. This technique was used to recover the aberrations of the Hubble Space Telescope [1,2] and has many potential applications, including the alignment of segmented-aperture telescopes such as the James Webb Space Telescope (JWST) [3] or the Thirty Meter Telescope. However, the application of focus-diverse phase retrieval, particularly for segmented-aperture systems, has been limited by the so-called "capture range" problem. If the starting guess is not close enough to the true phase, the algorithm may stagnate in a local minimum without reaching the true solution. In the initial alignment of a segmented system, the segment tips and tilts may be quite large, putting the system outside of the capture range of phase retrieval; this requires other methods to be employed until the segments are well enough aligned to allow phase retrieval to succeed [3–7].

Recent work by Thurman [8] has demonstrated an approximate phase retrieval method based on geometrical optics. This method can be used to obtain a rough estimate of the segment tips and tilts without itself requiring a good starting guess. This estimate can then serve as a starting guess for a more traditional phase retrieval algorithm, which will then be within its capture range. For a particular set of wavefront error statistics, Thurman's method was successful in only 60% of his trials. We will show some refinements of Thurman's method that can alleviate the capture range problem for segment tips and tilts in focus-diverse phase retrieval in all but a few cases.

Throughout this paper we will use a particular phase retrieval problem as an example to illustrate our techniques. Figure 1 shows the true wavefront error in this problem and its decomposition into individual components. The wavefront includes 0.064 wave RMS of intrasegment errors (11 Zernike-like terms) and 0.079 wave RMS of global aberration terms (11 Zernike-like terms). Intersegment errors consist of an additional 0.67 wave RMS tilt and 0.20 wave RMS piston. This falls roughly in the middle of the parameter range for the Monte Carlo study we will discuss in Section 5. The total wavefront error, excluding global tip/tilt and piston, is 0.75 wave RMS. If we additionally exclude wavefront components that can be represented as segment tip/tilt (including some of the global aberrations), the wavefront error is 0.24 wave RMS.

2. BASIC ESTIMATE PROCEDURE

The conceptual basis of Thurman's method is connecting the intensity measured in a particular PSF with the local ray density in that plane (as in a geometrical optics spot diagram) and with a probability density. To strengthen this connection, it is necessary to overcome the effects of diffraction by applying a blurring kernel to the detected PSFs:

$$p(x,y) \propto \mathcal{B} I_\ell(x,y), \tag{1}$$

where $p(x,y)$ denotes the probability of a ray passing through the given point, $\mathcal{B}$ is a blurring operator, I_ℓ is the measured PSF in the ℓth plane, and x,y are PSF-plane coordinates. For $\mathcal{B}$ we used a convolution with a Gaussian blurring kernel with a full width at half-maximum (FHWM) of $2\lambda f/\#$, where λ is the wavelength and $f/\#$ is the f-number of the system. Figure 2 shows the effect of this blurring on three of the five PSFs in our example problem. We can express any particular jth ray in the system in terms of the deviation from the ideal coordinates of that ray as follows. If x_j^{pup} is the x coordinate of

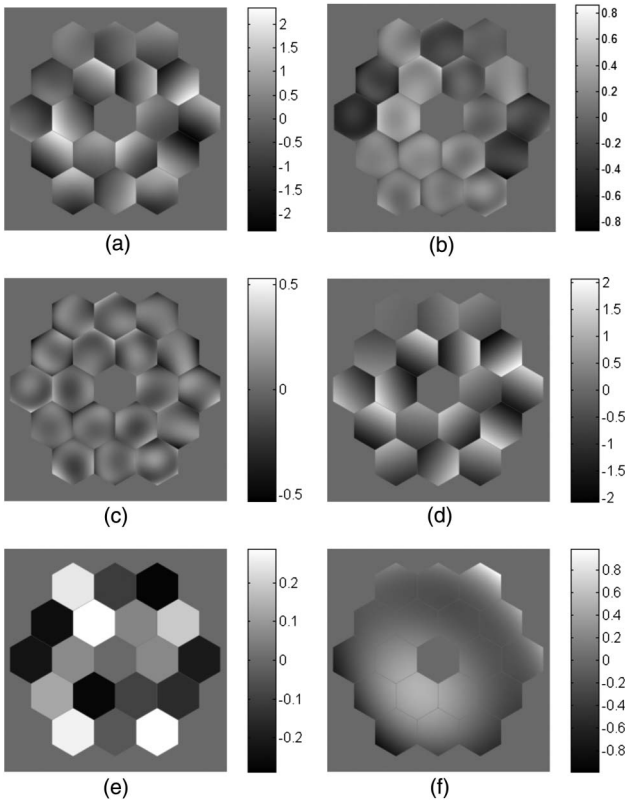

Fig. 1. Components of a representative wavefront error. (a) Total wavefront, (b) wavefront with the segment tip–tilt removed, (c) intra-segment aberrations, (d) segment tip–tilt component of the wavefront, (e) segment piston component, and (f) global component. Color bars are in units of waves for all figures.

that ray in the pupil, and the distance to the ℓth PSF plane is $f - \Delta z_\ell$, where f is the system focal length and Δz_ℓ is the defocus distance, then the x coordinate of the ideal ray is

$$x_{\ell j}^{\text{ref}} = \Delta z_\ell \frac{x_j^{\text{pup}}}{f} \tag{2}$$

and similarly for y. The actual ray coordinate in the ℓth plane is

$$x_{\ell j} = x'_{\ell j} + x_{\ell j}^{\text{ref}}, \tag{3}$$

where $x'_{\ell j}$ is the transverse displacement from the ideal ray, given paraxially by $x'_{\ell j} = \theta_{xj}(f - \Delta z_\ell) \approx \theta_{xj} f$, for small $|\Delta z_\ell|$, where θ_{xj} is the angular ray deviation from the ideal. Take the probability that the jth ray left the pupil with a given $(\theta_{xj}, \theta_{yj})$ and arrived at the ℓth PSF plane to be

$$p_{\ell j}(\theta_{xj}, \theta_{yj}) \propto \mathcal{B} I_\ell(\theta_{xj} f + x_{\ell j}^{\text{ref}}, \theta_{yj} f + y_{\ell j}^{\text{ref}}). \tag{4}$$

Considering the local ray density in one image plane as a probability suggests that we can also consider a joint probability density over all of the PSF planes as

$$p_j(\theta_x, \theta_y) \propto \prod_\ell p_{\ell j}(\theta_x, \theta_y). \tag{5}$$

We can make an estimate of the most probable ray deviations as

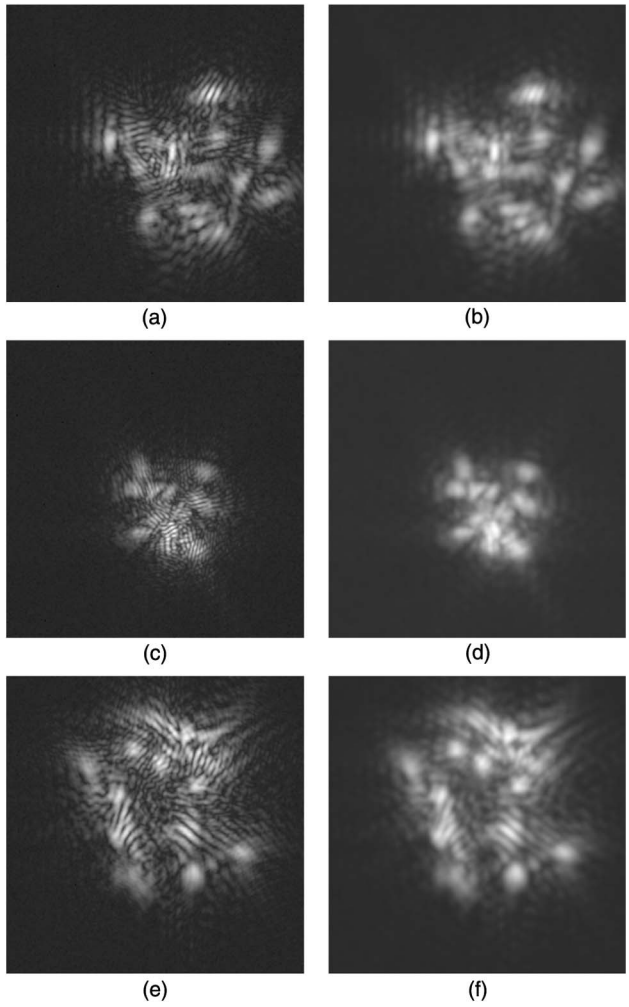

Fig. 2. Example PSFs with and without blurring. (a), (c), (e) PSFs without blurring, and (b), (d), (f) PSFs after blurring. PSFs are shown as intensity to the one quarter power. The top row is –8 waves out of focus, the middle row is nominally in focus, and the bottom row is +8 waves out of focus.

$$(\hat{\theta}_{xj}, \hat{\theta}_{yj}) = \arg \max_{(\theta_x, \theta_y)} [p_j(\theta_x, \theta_y)]. \tag{6}$$

From these angular deviations, we can directly obtain the local wavefront slopes (the wavefront slope angles are equal to the ray deviation angles) and through them the segment tips and tilts. We can think of the system in this case as being similar to a Hartmann sensor [9] (where the segments act as subapertures), in which the spots overlap one another and are confused; we use the multiple planes of data to disambiguate the Hartmann spots. Our simple way to visualize the process of determining the tip–tilt for a particular segment (a single ray) using this method is to imagine an observer sitting in the pupil plane at the location of the ray. The observer looks back through the PSF planes and chooses a ray direction so that it will pass through bright regions in all the PSF planes while avoiding dark regions.

3. ITERATIVE ESTIMATE PROCEDURE

Thurman's geometrical optics estimation technique described above may fail because the joint probability density for a given

segment may have more than one peak, suggesting multiple possible solutions for the tip–tilt of that segment. Furthermore, the peak with the largest value does not always correspond to the correct solution, because light originating from other segments confounds the analysis. Figure 3(a) shows a joint probability density of Eq. (5), with a clear single peak, whereas Fig. 3(b) shows one with multiple peaks. The wavefront in our example problem has several segments with joint densities having multiple peaks, three of which result in significant errors in the tip–tilt estimates. Figures 4(a)–4(c) show the tip–tilt estimates in this case. Figures 4(d) and 4(e) show the results of phase retrieval using the results in Fig. 4(b) as an initial estimate. Note that the three segments with the largest errors in the initial estimate have large residual errors after phase retrieval that was caught in a local minimum. To alleviate this problem, we developed a refinement to Thurman's method.

First, we define the set γ_k as those segments for which the joint probability density has only one significant peak (beginning with iteration number $k = 1$). Second, we form a mask, as shown in Fig. 5, which is unity for the set γ_k and zero for other (low-confidence) segments. We multiply estimated fields in the pupil by this mask and compute the partial PSFs, giving us an estimate of the light in each plane contributed by just those segments. Third, we subtract these partial PSFs from the measured PSFs to arrive at residual PSFs. Negative numbers in the residual PSFs are replaced by zero. Fourth, using the residual PSFs, we reestimate the tips and tilts for those segments in which we are not confident, denoted $\bar{\gamma}_k$, using the method described above (including the application of a blurring kernel). The residual PSFs take the role of the data. Figure 3(c) shows a resulting probability density for the residual PSFs; note that it has only a single peak as compared with the multiple peaks in Fig. 3(b).

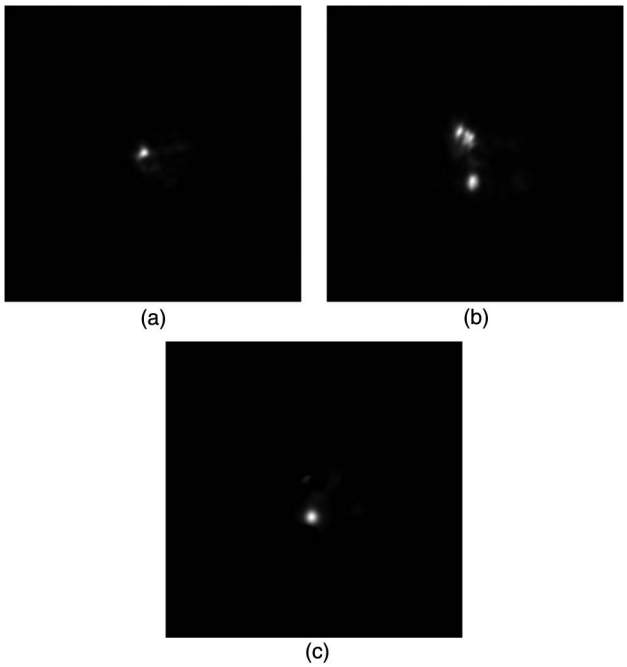

Fig. 3. Example probability densities (shown as the square root). (a) An unambiguous one with a single peak, (b) an ambiguous one with multiple peaks, and (c) the density from (b) after it has been corrected to remove ambiguity.

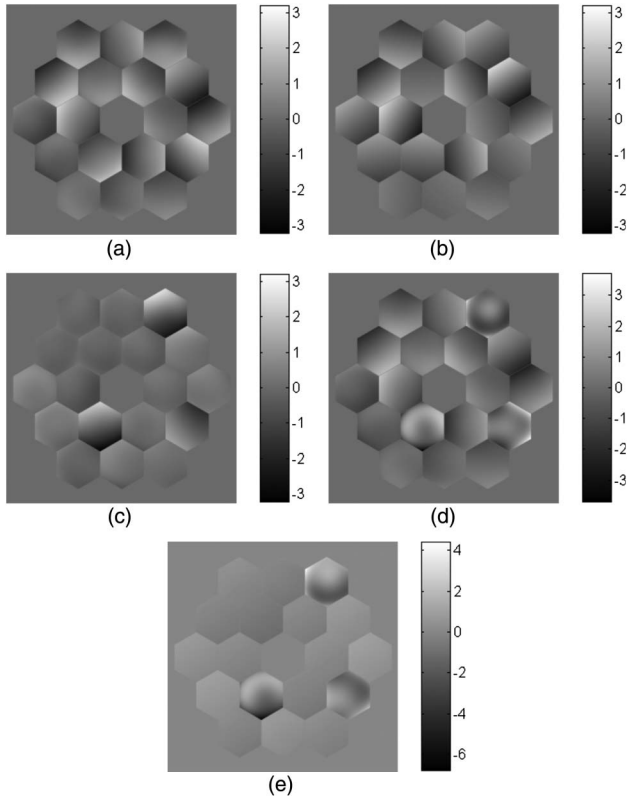

Fig. 4. Comparison of (a) a true wavefront, (b) the tip–tilt estimated by Thurman's method, (c) their difference, (d) the phase retrieval results starting from the initial ray-based estimate from Thurman's method, and (e) the error in the phase retrieval. Note that three particular segments have larger tilt errors in the ray estimate and higher order errors in the phase retrieval. The color bar units are waves.

We can then construct the new estimate as

$$\hat{\psi}_{k+1} = \begin{cases} \mathcal{K}_{\bar{\gamma}_k}[I - \mathcal{M}_{\gamma_k}[\hat{\psi}_{k,\gamma_k}]], & \bar{\gamma}_k \\ \hat{\psi}_{k,\gamma_k}, & \gamma_k \end{cases}, \tag{7}$$

where the set γ_k denotes the segments in which we are confident in our phase estimate on the kth iteration, $\bar{\gamma}$ denotes the segments in which we are not confident, I represents the measured data, $\hat{\psi}_{k,\gamma_k}$ denotes the phase estimate for the segments in the set γ_k only, and $\mathcal{M}_{\gamma_k}[\cdot]$ is a numerical forward model that uses the estimated pupil phases to form estimated PSFs for the segments in set γ_k. The procedure for applying $\mathcal{M}_{\gamma_k}[\cdot]$ is

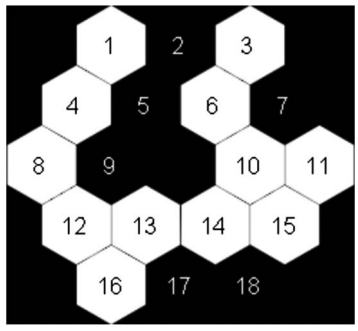

Fig. 5. For the examples, the mask identifying high-confidence segments used to remove light from the PSF and joint density functions.

as follows. Form a two-dimensional mask that is unity at every point within the segments that are part of γ_k and zero everywhere else. Form complex fields from this mask and the wavefront information. Propagate the fields to the nominal focus using numerical fast Fourier transform propagation, and from the nominal focus to each image plane using angular spectrum propagation (this model is appropriate for systems where the detector moves physically to introduce focus diversity). Take the squared magnitudes of the complex fields in the image planes to form image estimates; the result of $\mathcal{M}_{\gamma_k}[\cdot]$ is an entire focus-diverse intensity data set resulting from the provided wavefront estimates. The operator $\mathcal{K}_{\bar{\gamma}_k}[\cdot]$ denotes the process of using Thurman's method to construct a tip–tilt estimate for the segments in $\bar{\gamma}_k$. Equation (7) is repeated iteratively, with $\hat{\psi}_k$ being the phase estimate at iteration k.

In order to make progress, each iteration of Eq. (7) should identify at least one segment whose joint probability was previously ambiguous but that is unambiguous after removal of the light from the high-confidence segments. Thus, as we iterate, more segments are put into γ_k. Iteration continues until either all segments are in γ_k or until $\gamma_{k+1} = \gamma_k$ (meaning the algorithm can make no more progress). In this later case we take the highest probability peaks for the remaining segments and proceed with phase retrieval.

In the example problem, one iteration was sufficient to achieve an adequate tip–tilt estimate. Figure 6 shows the improved wavefront reconstruction in our example problem.

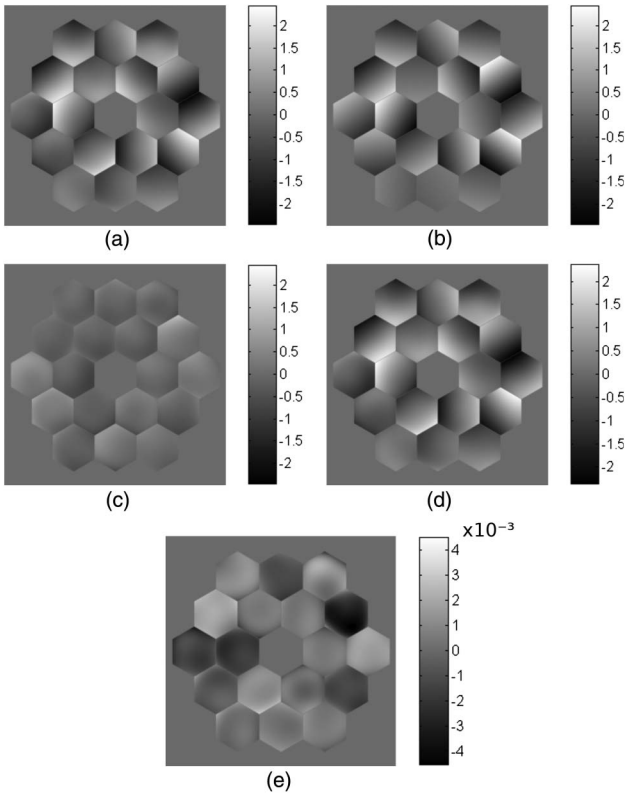

Fig. 6. Comparison of (a) the true wavefront, (b) the improved estimated wavefront by the ray-based estimate using Eq. (7), (c) their difference, which is improved from the result in Fig. 4(c), (d) phase retrieval results starting from the improved ray-based estimate, and (e) the error in phase retrieval. Note that the residual wavefront error in (e) is small (thousandths of a wave) for all segments. The color bar units are waves.

In most cases, this iterative process was sufficient to remove the ambiguity in the joint probability densities and arrive at good tip–tilt estimates.

4. AVOIDING PHASE RETRIEVAL STAGNATION

As in Thurman's previous work, we used the tip–tilt estimates from the geometrical optics method as a starting guess for phase retrieval. For the result shown in Fig. 4 we used a nonlinear optimization-based phase retrieval algorithm with a sum of squared differences of intensity error metric and a conjugate gradient search strategy using analytic gradients [1,10,11]. When we used Thurman's original algorithm, the starting guesses were close enough for phase retrieval to converge only 60% of the time. When we used our guesses derived from the improved algorithm described in Section 3, the rate of convergence was much higher, but still not 100%. Cases where the algorithm failed to find a good solution can be identified by comparing the final error metric value with the expected value based on the noise present in the data or by visually assessing the agreement between the modeled PSFs and the measured data. So typically the user of the algorithm would be aware that the fitting had failed in these cases rather than being misled to select an incorrect solution. We identified two primary modes for this stagnation. The first is tip–tilt errors in a single segment. Either as a result of an insufficiently accurate tip–tilt estimate in the starting guess or due to the stochastic nature of the nonlinear optimization in the phase retrieval algorithm, the tip–tilt for one of the segments may become significantly different from the correct value. This leads to stagnation in the phase retrieval algorithm, since generally moving energy from a segment having an incorrect tip–tilt estimate to its correct position would require the light to move away from a bright area through a dark area in one or more PSF planes to get to the correct bright area, raising the error metric. The gradient-based phase retrieval algorithm does not allow that. In this stagnated case, the phase retrieval algorithm was largely unable to fit the intrasegment aberrations as well as the tip–tilt of the segment having the incorrect tip–tilt, but generally the other segments achieved a reasonably good fit. This problem can be solved by repeating the geometrical optics analysis used to construct the tip–tilt starting guess for that segment, but taking into account the improved knowledge of the higher-order phase errors of the other segments. This requires first determining which segment is most likely to have an incorrect tip or tilt. We did this by considering the difference between the measured data and the modeled data:

$$\Delta I_\ell = I_\ell - \mathcal{M}_\ell[\hat{\psi}_{\mathrm{pr}}], \tag{8}$$

where $\hat{\psi}_{\mathrm{pr}}$ is the stagnated phase retrieval result and $\mathcal{M}_\ell$ is the forward model for the ℓth plane. In this case $\mathcal{M}_\ell$ does not include any masking but instead represents the model PSF including all segments. We blurred these differenced PSFs and converted them into angular space as in Eq. (4) to get

$$\varepsilon_{\ell j}(\theta_x, \theta_y) = \mathcal{B}\Delta I_\ell(\theta_x f + x^{\mathrm{ref}}_{\ell j}, \theta_y f + y^{\mathrm{ref}}_{\ell j}). \tag{9}$$

Note that it is no longer appropriate to consider $\varepsilon_{\ell j}$ as a probability density, even though it is similar in form to $p_{\ell j}$, since it

has a zero mean. Rather, $\varepsilon_{\ell j}$ represents the distribution of errors from the model in the ℓth plane as seen from the jth segment. We combined the information from the different planes, in this case using a sum rather than a product, given by

$$\varepsilon_j(\theta_x, \theta_y) = \sum_{\ell} \varepsilon_{\ell j}(\theta_x, \theta_y). \tag{10}$$

We used the peak-to-valley deviation in $\varepsilon_j(\theta_x, \theta_y)$ for the jth ray,

$$\Delta\varepsilon_j = \max[\varepsilon_j(\theta_x, \theta_y)] - \min[\varepsilon_j(\theta_x, \theta_y)], \tag{11}$$

as a good indicator of which segment is the most likely to have a wrong tip–tilt value. We selected the segment in error, j_e, by choosing the value of j for which $\Delta\varepsilon_j$ is largest. Given the stagnated phase retrieval result $\hat{\psi}_{\mathrm{pr}}$, we can form the new estimate,

$$\hat{\psi}_{\mathrm{new}} = \begin{cases} \mathcal{K}_{j_e}[I - \mathcal{M}_{\gamma}[\hat{\psi}_{\mathrm{pr},\gamma}]], & j_e \\ \hat{\psi}_{\mathrm{pr},\gamma}, & \gamma, \end{cases} \tag{12}$$

where γ is the set of all segments except the j_eth. Note that because $\hat{\psi}_{\mathrm{pr}}$ is the stagnated result of phase retrieval, it can also contain global phase terms; these should also be kept unchanged. Only the segment coefficients for segment j_e are changed. Specifically, the tip–tilt estimate of the j_eth segment should be updated; the higher-order segment coefficients of the j_eth segment are probably very wrong, and we set them to zero. Then phase retrieval is continued from this estimate. This approach is somewhat related to Zielinski and Fienup's approach in [12], but our approach can identify suspect segments without human assessment of PSFs and can iteratively improve the estimates without recourse to the full phase retrieval algorithm; Zielinski and Fienup's approach relies on reseeding suspect segments with random tip–tilt values.

The next common stagnation mode in this approach is an error in global tip–tilt if each PSF is allowed to have a separate global tip–tilt to account for camera motion or telescope jitter between frames. If these global estimates become incorrect, the phase retrieval algorithm can again stagnate. This is easily resolved by generating modeled PSFs using the stagnated phase retrieval result and registering each PSF against the simulated data for the same frame by finding the location of the peak of the cross correlation [13] between the modeled PSF and the data. That peak location is then used to compute a corresponding error in global tip–tilt, which is applied to the phase estimate. Phase retrieval is then restarted. This stagnation mode typically occurs as an error of several pixels in one of the focus-diverse image planes. It appears to be caused by a strong fringe in the part of the image being shifted over by at least one full period of the fringe, so that shifting the image to the correct position would require moving across the fringe nulls, temporarily increasing the error metric.

5. MONTE CARLO SIMULATION RESULTS

We applied the methods described above in a simulated system with an aperture consisting of 18 hexagonal segments with an average of 1.8 waves RMS of lower-order global aberrations, 0.4 wave RMS of lower-order segment aberrations, and segment tip–tilts ranging between zero and 1.4 waves RMS. All global and segment aberrations were assumed to be unknown. The system and simulation parameters are the same as those described in [8]. We used defocus settings of 0, ±6, and ±8 waves peak-to-valley, similar to but not the same as those planned for the JWST Near Infrared Camera (NIRCam). When the defocus is very large, the light from each segment forms a blob separate from the light from the other segments, as in a Hartmann test, making the identification of the segments easier. Including planes with larger defocus (such as the +12 waves for NIRCam [7]) would likely improve the performance of the basic method by reducing overlap of blobs from separate segments.

Applying the iterative tip–tilt estimation without the stagnation avoidance methods led to convergence in 183 of the 200 cases we considered (92% convergence success). By combining the improved starting guesses with the techniques for avoiding stagnation, we were able to achieve sucess in 196 cases (98% convergence success) for phase retrieval over the entire range of segment tip–tilts we considered. Of the additional 13 cases that converged with the stagnation avoidance techniques, 9 had stagnated due to single-segment tip–tilt errors and 4 had stagnated due to global tip–tilt errors.

Of the four trials that failed to converge after the application of stagnation avoidance, two represented cases where the first stagnation mode (tip–tilt error of a single segment) occurred and the tip–tilt reestimation process was not able to find a better solution. In the other two cases there were errors in the higher-order wavefront over the whole aperture, and the particular failure mode was not clear.

Figure 7 shows a comparison of the success rate between Thurman's method (dashed line), our improved method including the stagnation avoidance methods (solid line), iterative estimation without stagnation avoidance (solid line with squares), and phase retrieval used alone with a starting guess of zero phase (dotted–dashed line). Thurman's method works well with very large tip–tilts, where there is little chance for confusion between individual segments, and ordinary phase retrieval works well for small tip–tilts, where there is no need for tip–tilt bootstrapping. Our improvements help the most in the 0.25–1.25 wave RMS tip–tilt regime, where

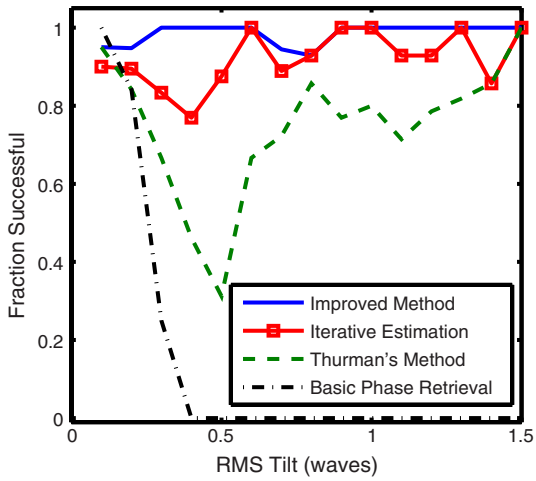

Fig. 7. Fraction of converged cases comparing the new method with iterative estimation, Thurman's approach, and standard phase retrieval. The cases are binned by RMS tip–tilt, with the upper edge of the bin interval on the horizontal axis.

ordinary phase retrieval algorithms are outside their capture range and the blobs in the PSF from different segments are not yet well enough separated for Thurman's algorithm to work reliably.

6. CONCLUSION

We have improved upon Thurman's geometrical optics tip–tilt estimation approach by introducing a method using the number of peaks within the joint probability density function to quantify our confidence in a particular estimate and use it to recursively reestimate tip–tilts, yielding more robust estimates. We also introduced a refinement of this approach that uses partially successful stagnated phase retrieval results to further improve the tip–tilt estimates. Finally, we introduced a method using image registration to avoid phase retrieval stagnation in global (per PSF) tip–tilt terms. Using these methods together allowed us to solve the capture range problem for segment tip–tilts; phase retrieval converged successfully in 98% of cases over a broad range of segment tip–tilt values.

ACKNOWLEDGMENTS

We thank Samuel Thurman for providing software for his method. This research was supported by funding from a University of Rochester Sproull Fellowship and from NASA Goddard Space Flight Center.

REFERENCES

1. J. R. Fienup, "Phase-retrieval algorithms for a complicated optical system," Appl. Opt. **32**, 1737–1746 (1993).
2. J. R. Fienup, J. C. Marron, T. J. Schulz, and J. H. Seldin, "Hubble Space Telescope characterized by using phase-retrieval algorithms," Appl. Opt. **32**, 1747–1767 (1993).
3. L. D. Feinberg, B. H. Dean, D. L. Aronstein, C. W. Bowers, W. Hayden, R. G. Lyon, R. Shiri, J. S. Smith, D. S. Acton, L. Carey, A. Contos, E. Sabatke, J. Schwenker, D. Shields, T. Towell, F. Shi, and L. Meza, "TRL-6 for JWST wavefront sensing and control," Proc. SPIE **6687**, 668708 (2007).
4. D. S. Acton, P. D. Atcheson, M. Cermak, L. K. Kingsbury, F. Shi, and D. C. Redding, "James Webb Space Telescope wavefront sensing and control algorithms," Proc. SPIE **5487**, 887–896 (2004).
5. D. S. Acton, T. Towell, J. Schwenker, J. Swensen, D. Shields, E. Sabatke, L. Klingemann, A. R. Contos, B. Bauer, K. Hansen, P. D. Atcheson, D. Redding, F. Shi, S. Basinger, B. Dean, and L. Burns, "Demonstration of the James Webb Space Telescope commissioning on the JWST testbed telescope," Proc. SPIE **6265**, 62650R (2006).
6. D. S. Acton, T. Towell, J. Schwenker, D. Shields, E. Sabatke, A. R. Contos, K. Hansen, F. Shi, B. Dean, and S. Smith, "End-to-end commissioning demonstration of the James Webb Space Telescope," Proc. SPIE **6687**, 668706 (2007).
7. D. S. Acton, J. S. Knight, A. Contos, S. Grimaldi, J. Terry, P. Lightsey, A. Barto, B. League, B. Dean, J. S. Smith, C. Bowers, D. Aronstein, L. Feinberg, W. Hayden, T. Comeau, R. Soummer, E. Elliott, M. Perrin, and C. W. Starr, "Wavefront sensing and controls for the James Webb Space Telescope," Proc. SPIE **8442**, 84422H (2012).
8. S. T. Thurman, "Method of obtaining wavefront slope data from through-focus point spread function measurements," J. Opt. Soc. Am. A **28**, 1–7 (2011).
9. D. Malacara, *Optical Shop Testing* (Wiley, 2007).
10. J. R. Fienup, "Phase retrieval algorithms: a comparison," Appl. Opt. **21**, 2758–2769 (1982).
11. S. T. Thurman and J. R. Fienup, "Phase retrieval with signal bias," J. Opt. Soc. Am. A **26**, 1008–1014 (2009).
12. T. Zielinski and J. R. Fienup, "Extending wavefront sensing capture range for segmented systems through tip and tilt estimation," in *Frontiers in Optics* (Optical Society of America, 2006), paper FMF4.
13. M. Guizar-Sicairos, S. T. Thurman, and J. R. Fienup, "Efficient subpixel image registration algorithms," Opt. Lett. **33**, 156–158 (2008).