Machine learning for improving stellar image-based alignment in wide-field Telescopes

Zhi-Xu Wu¹, Yi-Ming Zhang¹, Rong-Xin Tang¹, Zheng-Yang Li², Kai-Yuan Zhang², Xiang-Yan Yuan², Yong Xia¹, Hua Bai², Bo Li², Zhou Chen¹, Xiang-Qun Cui² and Xiao-Hua Deng¹

¹ Institute of Space Science and Technology, Nanchang University, Nanchang 330031, China; ² Nanjing Institute of Astronomical Optics and Technology, Chinese Academy of Sciences, Nanjing 210042, China;  *Corresponding author: zyli@niaot.ac.cn, xyyuan@niaot.ac.cn

Received 2021 August 5; accepted 2021 October 25

Abstract The stellar images will deteriorate dramatically when the sensitive elements of the wide-field survey telescopes are misaligned during the observation, and active alignment is the key technology to maintain the high resolution of the wide-field sky survey telescopes. Instead of traditional active alignment based on the field-dependent wavefront errors, this work proposed a machine learning alignment metrology based on stellar images of the scientific camera, which is more convenient and high speed. We first theoretically analyzed that the pattern of the PSF over the field is closely related to the misalignment status, and then the relationships are learned by the two-step neural networks. After two-step active alignment, the position errors of misalignment parameters are over 90% below 5µm and over 90% angle positions below 5 arcsec. The precise alignment results indicate that this metrology provides a low-cost and high-speed solution to maintain the image quality of the wide-field sky survey telescopes during observation, thus has important significance and broad application prospects.

Key words: techniques: high angular resolution — techniques: active alignment — techniques: machine learning — telescopes

1 INTRODUCTION

In modern astronomy, such as time domain astronomy, black hole, dark energy, etc. (Kasliwal et al. 2019; Copeland et al. 2006; Yu 2020), a large field of view and high resolution is presumed for the sky survey telescope to accelerate the discoveries in astronomy and astrophysics. As the field and the aperture of the sky survey telescopes increases, the primary mirror with a fast focal ratio is preferred to reduce the obscuration of the secondary mirror and decrease the length of the barrel, such as the J-PAS has a primary focal ratio of 1.5 (Benitez et al. 2014), the primary focal ratio of LSST is 1.18 (Sebag et al. 2016), and Mephisto is 1.3 (Yuan et al. 2020; Li et al. 2020). However, along with the fast focal ratio of the primary mirror, the second mirror will be more sensitive to the misalignment and the structural deformation caused by gravity and temperature should be smaller.
The field-dependent aberration behaviors of a misaligned telescope with two or three mirrors were demonstrated by many works, Baranne, and Wetherell et al. have derived the formula of the coma aberration for the non-coaxial dual-mirror system in 1972 (Baranne 1966; Wetherell & Rimmer 1972). For the more common situation, Dingqiang Su has analyzed the coma aberration of the non-coaxial dual-mirror system when the secondary mirror with decent and tip-tilt and the two axials of the secondary mirror and primary mirror are not lie in the same plan, the coma aberrations are also proved that they can be operated like vectors (Su 1989). According to the Nodal Aberration Theory, which is first reported by Shack and Thompson (Shack & Thompson 1980; Thompson et al. 2008), the orthogonal Zernike polynomials are widely used for numerical expression of field-dependent aberrations, algorithms such as reverse-optimization, damped least-square (Lee et al. 2007; Bloemhof et al. 2012; Li et al. 2015b,a), and principal components analysis, can deliver suitable corrections for resolving misalignments with relatively high accuracy.

Unfortunately, acquisitions of field aberrations are difficult during telescope observations and complex wavefront sensing systems have to be equipped for measuring the field aberrations. Four edge field curvature wavefront sensors at the focal plane are the most commonly used for wide-field imagers with high image quality, such as the Javalambre Survey Telescope (JST/T250) and Large Synoptic Survey Telescope (LSST) (Chueca et al. 2012; Manuel et al. 2010; Xin et al. 2015; Claver et al. 2012). However, a beam splitting system, which is required for the traditional curvature wavefront sensor to measure the in-focus and ex-focus images, is easy to produce vignetting due to the fast focal ratio. The split curvature wavefront sensors need to compensate for the inherent aberrations of the two sources with different field-of-view and require expensive CCD stitching technology, which significantly increases the cost of the camera.

Instead of detecting the field wavefront errors, the pattern of the PSF ellipticity distribution also revealing the system’s misaligned status (K. et al. 2016), and measuring the PSF of the scientific images seems to be more convenient (Luppino & Kaiser 1997). In this work, a machine learning alignment metrology is proposed for calculating the misalignments parameters based on the stellar images of the wide-field Telescopes, and we employ a relatively complex R-C configuration (Mephisto) to model the alignment metrology. The optical system is perturbed with various sets of misalignments delivering misaligned field-dependent PSF, and the explicitly mathematical model between misalignments and the pattern of field-dependent PSF is complex and hard to establish. Machine learning is a method of data analysis that automates analytical model building (Chen et al. 2019; Tang et al. 2020), and the nonlinear nature of neural networks along with their inherent flexibility and adaptability makes them good candidates for stellar image-based alignment that are not easily solved with conventional algorithms. Compared with the traditional active alignment metrology based on field-dependent aberrations, there are some advantages. Firstly, the active alignment system can be simplified because the wavefront sensor system has been removed. Secondly, as the one-to-one correspondence between the stellar images of the CCD and the output misalignment parameters, the calculation time spent on the wavefront reconstruction and misalignment parameters calculation form field depended wavefront error can be avoided, which improves the response speed of the active alignment.

2 METHOD

2.1 the pattern variations of the PSF resulting from misalignment

Generally, the stellar image is a convolution of an ideal point source and the PSF. The wavefront expression for the misaligned optical system is given by Eq(1) expressed as:

\[
W = \sum_{j} \sum_{p=0}^{\infty} \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} (W_{klm}) [(H - \sigma_j) \bullet (H - \sigma_j)]^p \times (\rho \bullet \rho)^n [(H - \sigma_j) \bullet \rho]^m
\]  
(1)
where $W_{klm}$ is the aberration coefficient, $H$ represents the normalized image field position, $\rho$ is the normalized pupil position and $\sigma_j$ denotes the aberration field center shifting vector of number $j$ elements, additionally, $k=2p+m, l=2n+m$. If $i$ represents the total misaligned degree of freedoms of a misaligned system, for any field of the optical system, the wavefront aberration at the pupil is obtained from Eq (2) given as:

$$W = W(\delta_1, \delta_2, \ldots, \delta_i, \rho)$$  \hspace{1cm} (2)

where $p(\rho)$ is the pupil function, and the generalized pupil function $P(\delta_1, \delta_2, \ldots, \delta_i, \rho)$ is therefore given by Eq (3).

$$P(\delta_1, \delta_2, \ldots, \delta_i, \rho) = p(\rho) \exp jW(\delta_1, \delta_2, \ldots, \delta_i, \rho)$$  \hspace{1cm} (3)

The amplitude spread function $h(H)$ is a Fourier transfer of the generalized pupil function shown in Eq (4).

$$h(H) = FFT(P(\delta_1, \delta_2, \ldots, \delta_i, \rho))$$  \hspace{1cm} (4)

$$PSF(\delta_1, \delta_2, \ldots, \delta_i, H) = |h(\delta_1, \delta_2, \ldots, \delta_i, H)|^2$$  \hspace{1cm} (5)

Fig. 1: PSF representation in the image plane for the situation with or without misalignment. The PSF image on the left shows a standard airy disk formed by an ideal optical system; the right image is a PSF of an optical system with coma aberration.

The PSF distribution in the image plane is related to the misalignments $(\delta_1, \delta_2, \ldots, \delta_i)$. To numerically demonstrate the PSF distribution, a measurable parameter termed ellipticity is adopted. This term was previously used in weak lensing research for describing the shape of galaxies (Luppino & Kaiser 1997).

$$\left\{ \begin{array}{l} Q_{i,j} = \int_{(H_x,H_y)} weight(H_x,H_y) \text{PSF}(\delta_1, \delta_2, \ldots, \delta_i, H_x, H_y) H_x H_y dH_x dH_y \\ e_\alpha = \frac{Q_\alpha}{T} \end{array} \right.$$  \hspace{1cm} (6)

In Eq(6), $\alpha = 1 \ or \ 2, H_1 = H_x - H_{x\text{center}}, H_2 = H_y - H_{y\text{center}}, Q_1 = Q_{11} - Q_{22}, T = Q_{11} + Q_{22}$. Hence, with PSF acquisition at a field of view $(H_{x0}, H_{y0})$, the numerical ellipticity is expressed as:

$$\left\{ \begin{array}{l} E_{10} = e_1(\delta_1, \delta_2, \ldots, \delta_i, H_{x0}, H_{y0}) \\ E_{20} = e_2(\delta_1, \delta_2, \ldots, \delta_i, H_{x0}, H_{y0}) \end{array} \right.$$  \hspace{1cm} (7)
In addition to ellipticity, FWHM (Full width at half maximum) R and Azimuth $\theta$ are also the parameters that can numerically describe the PSF.

$$\begin{align*}
\epsilon &= \frac{Q_{ii} - Q_{jj} + 2iQ_{ij}}{Q_{ii} + Q_{jj}} \\
R &= \sqrt{Q_{ii} + Q_{jj}} \\
\theta &= \tan^{-1}\frac{Q_{ij}}{Q_{ii} - Q_{jj}}
\end{align*}$$

(8)

To illustrate the field-dependent PSF variations resulting from misalignment, we deliberately perturbed the secondary mirror of Mephisto and recorded $e_1$ and $e_2$ at diagonal fields of view. Figure(3) A and B show that $e_1$ and $e_2$ vary from a FOV of (-0.7°, -0.7°) to a FOV of (0.7°, 0.7°) with secondary mirror decentre values of -0.04, -0.02, 0.02, and 0.04mm at the sagittal surface; Figure(3) C and D show $e_1$ and $e_2$ vary from a FOV of (-0.7°, -0.7°) to a FOV of (0.7°, 0.7°) with secondary mirror decentre value of -0.04, -0.02, 0.02, and 0.04mm at the meridian surface.

![Fig. 2: PSF ellipticity $e_1$ and $e_2$ over the diagonal field of view of the nominal design. The range of the diagonal FOV is from (-0.7°, -0.7°) to (0.7°, 0.7°), the distributions pattern of $e_1$ and $e_2$ are symmetric distributed.](image-url)
Fig. 3: PSF ellipticity $e_1$ and $e_2$ with decentring $(X, Y)$ of the secondary mirror. The range of the diagonal FOV is from $(-0.7^\circ, -0.7^\circ)$ to $(0.7^\circ, 0.7^\circ)$. A and B show the distributions patterns of $e_1$ and $e_2$ when the secondary mirror of the optical system with x-decent misalignment; C and D are the $e_1$ and $e_2$ patterns of the secondary mirror with y-decent. It’s shown that the symmetric features no longer exist due to the decent misalignment of the secondary mirror.

Fig. 4: PSF ellipticity $e_1$ and $e_2$ with tilt $(X, Y)$ of the secondary mirror. The range of the diagonal FOV is from $(-0.7^\circ, -0.7^\circ)$ to $(0.7^\circ, 0.7^\circ)$. A and B show the distributions patterns of $e_1$ and $e_2$ when the secondary mirror of the optical system with x-tilt misalignment; C and D are the $e_1$ and $e_2$ patterns of the secondary mirror with y-tilt. The symmetric features are also disappeared.
The perturbation parameters curves reveal that: (1) misalignment destroys the symmetry of ellipticity distribution; (2) over the field of view, the ellipticity variations are continued; (3) the secondary decentre in the sagittal surface may couple in tilt in the meridian surface; (4) the ellipticity distributions vary non-linearly with the misalignments. Based on the observations above, calculating the misalignments by mapping the PSF distribution of a scientific image is feasible.

2.2 Machine learning for active alignment

2.2.1 Generation of the training dataset

Machine learning is essentially a data-driven method, that requires a large amount of field-dependent PSF and corresponding misalignment parameters to feed the neural networks. Because different telescopes have different optical configurations, a deep learning model trained with one particular telescope cannot be adjusted and transferred to another telescope. To obtain the neural network model for one specific telescope, a customized training dataset of Mephisto is necessary. The Mephisto is developed for Yunnan University, it combines a large aperture with a wide field of view for rapid survey while ensuring an optical imaging quality suitable for precise photometry. The innovative optical design guarantees an ambitious project enabling the observations of three colors simultaneously, with wavelengths ranging from 320nm to 1000nm. The specifications of Mephisto are presented in Table 1.

Table 1: Specifications for Mephisto including optics, aperture, focal ratio, the field of view, pixel scale, and image quality

<table>
<thead>
<tr>
<th>ITEM</th>
<th>Specifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optics</td>
<td>Ritchey-Chrétien with field corrector</td>
</tr>
<tr>
<td>Aperture</td>
<td>1.6m (diameter)</td>
</tr>
<tr>
<td>Focal ratio</td>
<td>4.5</td>
</tr>
<tr>
<td>Field of View</td>
<td>2 degrees (diameter)</td>
</tr>
<tr>
<td>Pixel scale</td>
<td>0.286 arcsec/10um</td>
</tr>
<tr>
<td>Image quality</td>
<td>80% EE ≤ 0.6 arcsec</td>
</tr>
</tbody>
</table>

In Mephisto, there are 8 degrees of freedom (DOF) perturbations, which including 5 rigid body DOFs of M2 and 3 rigid body DOFs of the camera. The hexapod microrobots are applied on the M2 and the camera to compensate for the rigid body misalignments. The acquisition of training data is automatically executed by Dynamic Data Exchange (DDE) between MATLAB and ZEMAX. Firstly, 8 DOFs perturbations with different sensitivity are generated randomly through MATLAB. Secondly, these misaligned parameters are set into ZEMAX by DDE programming. Thirdly, the focal plane is partitioned into $m \times m$ equal tiles, PSF in the cent of each tile is generated by ray tracing in ZEMAX and imported into MATLAB, and this PSF is described as measurable parameters. Assume that the misalignment state of Mephisto is $s_I$, and $3 \times m^2$ measurable parameters of the PSF can be expressed as:

$$ (e_{s_1,f_1}, e_{s_1,f_2}, \ldots, e_{s_m,f_m}, \ldots, e_{s_1,f_{m^2}}) \quad (9) $$

$$ (\theta_{s_1,f_1}, \theta_{s_1,f_2}, \ldots, \theta_{s_m,f_m}, \ldots, \theta_{s_1,f_{m^2}}) \quad (10) $$

$$ (r_{s_1,f_1}, r_{s_1,f_2}, \ldots, r_{s_m,f_m}, \ldots, r_{s_1,f_{m^2}}) \quad (11) $$

Because the stellar images are distributed randomly on the CCD, the random positions of the stellar images will significantly increase the learning complexity and scale of deep learning. In a coaxial optical system, the pattern of the field-depend PSF is generally distributed axially symmetrical, which means that the pattern is distributed symmetrically around the center point,
and the symmetrical pattern will be broken when the co-axial optical system is misalignment. This phenomenon is very similar to field-dependent aberrations. Analogous to the vector aberration theory, the $3 \cdot m^2$ measurable parameters are also fitted by low-order Zernike polynomial to $3 \cdot 10$ parameters in this work.

Fig. 5: Layout of the misalignment perturbations of Mephisto. The 5 rigid body DOFs of M2 contain x-decent, y-decent, x-tilt, y-tilt, and piston; the 3 rigid body DOFs of the camera are x-tilt, y-tilt, and piston.

2.2.2 Network architecture

To improve the training efficiency, two-step calibration by coarse and fine neural networks are used in this work. The telescope is calibrated firstly by the coarse neural network, and then the hexapod microrobots compensate for the rigid body motions according to the parameters outputted by the fine neural network. The coarse network has the identical architecture as the fine network, while with different training ranges. The network architecture is shown in Figure(6), each hidden layer contains 300 nodes, the input layer has 30 input nodes for low-order Zernike polynomial coefficients, and the output layer contains 8 output nodes for the misalignment parameters.
Fig. 6: The network architecture of the coarse neural network and fine neural network. The architecture of the two networks is identical. The nodes of the input layer are 30; each hidden layer contains 300 nodes; the output layer contains 8 nodes.

Since the three measurable parameters of the PSF are in different ranges and the sensitivities of each misalignment parameter are different, normalization is necessary to ensure that all input vectors and the output misalignment parameters are within the same scale. It’s noticeable that the original scale of the input training data is saved, thus the normalization will not affect the relationship between the pattern of the PSF over the field and the misalignment status. The normalization is expressed as:

$$X_{\text{scaled}} = \frac{X - X_{\text{mean}}}{X_{\text{std}}}$$  \hspace{1cm} (12)

where $X_{\text{scaled}}$ is the normalized dataset, $X$ is the original dataset, $X_{\text{mean}}$ is the mean value of the original dataset, $X_{\text{std}}$ is the standard deviation of the original dataset.

Neural networks are trained using stochastic gradient descent of the loss function. Two kinds of loss function, MAE loss function, and Log-cosh loss function are used in this work. The stochastic gradient descent of the loss function can be expressed as:

$$\begin{align*}
   w_{ij}^{L} &= w_{ij}^{L} - \alpha \frac{\partial}{\partial w_{ij}^{L}} \text{Loss}_{\text{Fun}} \\
   b_{i}^{L} &= b_{i}^{L} - \alpha \frac{\partial}{\partial b_{i}^{L}} \text{Loss}_{\text{Fun}}
\end{align*}$$  \hspace{1cm} (13)

$$\text{Loss}_{\text{Fun}} = \begin{cases} 
   \text{MAE} \left( \text{Net}_{i}^{In}, \text{Net}_{i}^{Out} \right) = \frac{1}{n} \sum_{i=1}^{n} |\text{Net}_{i}^{In} - \text{Net}_{i}^{Out}| \\
   \text{LogCosh} \left( \text{Net}_{i}^{In}, \text{Net}_{i}^{Out} \right) = \sum_{i=1}^{n} \log (\cosh \frac{\text{Net}_{i}^{In} - \text{Net}_{i}^{Out}}{n})
\end{cases}$$  \hspace{1cm} (14)

where $\text{Net}_{i}^{In}$ is the label of the misalignment parameters, and $\text{Net}_{i}^{Out}$ is the output data of the training dataset. Because accuracy rate is easier than loss function to interpret and monitor during the training phase, we define the accuracy rate as the evaluation of the final model accuracy:

$$\text{Acc} = \frac{m}{n} \times 100\%$$  \hspace{1cm} (15)

where $m$ is the number of the predicted value, which satisfy the requirements in the coarse neural network and fine neural network. $n$ is the sample size of the test dataset.
3 RESULT

3.1 Configuration of the simulation

In our simulations, the adopted neural network was trained on a laptop with an Intel i7 9750h processor, 16 GB RAM, and an Nvidia GTX 1660ti laptop graphics card with 6 GB VRAM. TensorFlow, a widely used Python deep learning library, was utilized in neural network training. We randomly split the 10000 stellar images and the corresponding misalignment parameters acquired from the DDE programming to 70%, 20%, and 10% as training, validation, and testing data, respectively. Note that the measurable parameters are fitted by the low-order Zernike polynomials to simplify the training model, the model can be trained successfully with a laptop and the prediction of the misalignment parameters of the two-step neural networks is almost real-time. The parameters (learning rate, loss function, active function) are updated in the networks for 400 epochs by using Optuna, which is an automatic hyperparameter optimization software framework, and it features an imperative, define-by-run style user API (Akiba et al. 2019). By using dropout to prevent Neural Networks from overfitting, loss function can convergence after training 400 epochs as shown in fig(7).

![Loss Curve](image)

Fig. 7: Loss function curve of the neural network. The loss function is converged after 400 epochs of training.
Fig. 8: PSF, parameters of PSF and Zernike coefficients before correction. The right column is the PSF images of 25 fields of view; the middle column shows the calculated three parameters of the 25 PSFs, and the three parameters are ellipticity, radius, and theta; the left column is the fitted Zernike coefficients of the three parameters in the middle column. It’s noticeable that all parameters have no units due to the normalization.

Fig. 9: PSF, parameters of PSF and Zernike coefficients after correction. It’s shown that the performance of the system has improved greatly after the stellar image-based alignment.
3.2 Simulation results

In the simulation of the training dataset, the range of the position is (-1mm, 1mm) and the range of the tilt-tip angle is (-5 arcmin, 5 arcmin). The stellar images are generated by ray tracing with random misalignment parameters in the range, which are shown in the right column of Fig(8). The three measurable parameters of the PSF are shown in the middle column of Fig(8), and then fitted to 10 low-order Zernike polynomials are shown in the left of Fig(8). The fitted low-order Zernike polynomials are the input and the corresponding label are the misalignment parameters. During training, the loss function is converged quickly, which is shown in Fig(7). After two-step active alignment by using a coarse neural network and fine neural network, the telescope can be calibrated precisely. The stellar images of the calibrated telescope are shown in Fig(9). The spot diagram is very small except for the four PSF at the corners, which are out of the working field. This phenomenon is more obvious in Fig(10), for the 1000 test dataset, the RMS of the spot diagram of 1000 testing datasets at five fields of views are distributed randomly in the range of 50-200um. After two-step active alignment, the RMS is mainly below 18um. The error of the misalignment parameters distribution is shown in Fig(12). The first five pictures are the misalignment parameters of the second mirror, the last three pictures are the misalignment parameters for the focal plane. The errors are distributed like Gaussian function, over 90% position errors below 5um and over 90% angle positions below 5 arcsec. The results showed that the metrology proposed in this work is very accurate to actively align the telescope.
Fig. 10: RMS of the spot diagram before correction. The five subfigures describe the distribution of PSF’s RMS of 1000 testing samples at five FOV. These five FOVs contain four corner FOV of (1°, 0°), (0°, 1°), (-1°, 0°), (0°, -1°), and one center FOV of (0°, 0°). The RMS of the spot diagram are distributed randomly in the range of 50-200µm.
Fig. 11: RMS of the spot diagram after correction. The five subfigures describe the distribution of PSF’s RMS of 1000 testing samples at five FOV. These five FOVs contain four corner FOV of (1°, 0°), (0°, 1°), (-1°, 0°), (0°, -1°), and one center FOV of (0°, 0°). After two-step active alignment, the RMS are mainly below 18um.
4 CONCLUSION

A novel two-step active alignment metrology based on the coarse and fine neural networks is proposed in this work. The model calculating the misalignments based on stellar images of the wide-field Telescopes, and the two-step neural network can output the misalignment parameters of the optical system directly. Once it is well trained. Compared to the conventional field-dependent aberration approaches, this method is simpler in the system and higher in the calculating speed. Simulations are implemented to determine that the two-step neural network can be trained to calculate the misalignment parameters efficiently and accurately. Currently, the two-step neural network does not consider gravitational deformation of the primary mirror. Powerful learning ability can be used to continue predicting the misalignments including the deformation of the primary mirror for better applicability. Future work is considered to generate the training dataset through the experiment system, which is used to actively align the misalignments of the Mephisto. This work presents a feasible and easy-implemented method to improve the efficiency and accuracy of the active alignment metrology. Besides laboratory alignment, the metrology proposed in this study aims to maintain image quality during observation, with relatively low misalignment values mostly resulting from gravitational deformation. This method also has broad application prospects in the off-axis system and the optical system which contain freeform surface.

Acknowledgements This work was supported by the National Natural Science Foundation of China(Grants No. U1931207, 12173062). The authors acknowledge Dr. Zhou Chen at Institute of Space Science and Technology for helpful discussions on artificial network designs.

References
