IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 4, 2020, accepted June 10, 2020, date of publication June 16, 2020, date of current version June 26, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3002901

Extending Capture Range for Piston Error in Segmented Primary Mirror Telescopes Based on Wavelet Support Vector Machine With Improved Particle Swarm Optimization

HAIFENG CAO[1,2], JINGXU ZHANG[2], FEI YANG[2], QICHANG AN[2], AND YE WANG[1,2]

[1]Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China
[2]College of Materials Science and Opto-Electronic Technology, University of Chinese Academy of Sciences, Beijing 10049, China

Corresponding author: Fei Yang (yangflying@163.com)

This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFE0102900, in part by the Youth Innovation Promotion Association CAS under Grant 2016198, in part by NSFC under Grant 11673080 and Grant 11973040, and in part by the Jilin Science and Technology Development Program under Grant 20180520171JH.

ABSTRACT In the co-phasing techniques applied to the segmented telescope, a Shack–Hartmann wavefront sensor cannot accurately detect the piston error of the segment. Although the phase diversity (PD) algorithm can detect the piston error of each segment, it fails to reconstruct the wavefront quickly, and its dynamic range is small. Other technologies, such as prisms or micro-lens arrays, will significantly increase the complexity and construction cost of the optical system. Moreover, they may also introduce non-common path errors. In this study, we propose an approach to address this challenging problem via curvature sensing. This method uses multi-wavelength to eliminate the influence of 2π ambiguity and improve the capture range of co-phasing detection. However, curvature sensing is easily influenced by atmospheric seeing. We propose a wavelet support vector machine optimized via particle swarm optimization (PSO-WSVM) method to deal with this problem, and to improve the application scope of curvature sensing. We reshape SVM with a wavelet kernel function, and improve the PSO algorithm. We train the SVM to build a prediction model to distinguish the piston error range of each pair of adjacent segments and surpass 2π ambiguity. First, we obtain defocused images by means of the convolution technique. Second, we propose a prediction model based on SVM. We select the correlation coefficient between the sampling signal and the template signal at different wavelengths as the input vector, and we choose a wavelet basis function as the kernel function of SVM. Third, we improve the PSO algorithm with the exponential decreasing inertia weight (EDIW) to tune the parameters of SVM. Finally, we perform a simulation experiment on a real optical system model based on the Keck telescope. The results indicate that the performance of this method is better than that of other state-of-the-art SVM-based classifiers, and it works rapidly during the observation.

INDEX TERMS Image analysis, optical sensors, machine learning, algorithms, active optics, mirrors, phase measurement, PSO, support vector machine (SVM).

I. INTRODUCTION

Astronomy is about discovery. Astronomers need larger and larger telescopes to observe more distant and weaker celestial phenomena across vast distances and enormous spans of time. However, segmentation is a necessity to build telescopes larger than the 8 m class due to limitations in processing, manufacturing, transportation, and cost. A larger segmented mirror means a stronger ability to collect light, and a higher spatial resolution, but the discontinuous primary mirror has difficulty co-phasing due to gravity, wind, thermal drift, movement during slewing and other factors. In order to achieve the same optical performance as a monolithic mirror, the piston errors between the segments should be controlled to an accuracy of several nanometers.

The associate editor coordinating the review of this manuscript and approving it for publication was Shuihua Wang.

The methods used for co-phasing segmented mirrors can be divided into three categories based on the measurement principle as follows: (1) Image plane techniques, which mainly include phase retrieval (PR) and phase diversity (PD). (2) Pupil plane techniques, which include modified Shack-Hartmann sensing, the pyramid sensor, and the Mach-Zehnder Interferometer. (3) Intermediate plane techniques, which generally refer to curvature sensing (CS) [1]–[6].

The PR and PD techniques record the light intensity distribution on the specific plane of the charge-coupled device (CCD) existing on the telescope and reconstruct the wavefront by using the recorded light intensity information [7], [8]. PR only records the intensity distribution on the image plane. The wavefront reconstructed using an iterative algorithm is not unique and can only be used when the object is a point source. PD uses both light intensity distribution information on the image and defocused planes to reconstruct the wavefront. Unlike the PR method this method reconstructs the wavefront phase to be unique and can be used for the wavefront phase reconstruction of a light source object with a specific shape. The advantage of this technique is that the device is simple, whereas the disadvantage is that the measurement range of the phase is small ($\pm\lambda_{used}/2$) and it is highly computationally intensive. Moreover, it is susceptible to atmospheric disturbances.

Shack-Hartmann sensing has the advantages of a high utilization rate of light energy, broad dynamic range, low environmental requirements, and suitability for detection with white light. This method has been successfully applied in the phase camera system (PCS) of the Keck Telescope and the alignment and phasing system (APS) of the Thirty Meter Telescope (TMT). Pyramid sensing has an adjustable sensitivity and sample rate that can be used to detect higher-order aberrations, but the sensors require an enormously high manufacturing accuracy. These methods require micro-lens or prism arrays with fine tolerance over the edges of each pair of adjacent mirror segments. There are 36, 492, and 984 segments for the Keck, TMT, and the European Extremely Large Telescope, respectively. Therefore, these methods are becoming progressively less practical for future large telescopes. Other methods based on dispersed fringe sensors, point spread function (PSF) method or modulation transfer function (MTF) need to measure piston error in another optical path, which will introduce non-common path errors and be easily affected by the atmosphere [9]–[11].

In 1988, F. Roddier proposed CS. The wavefront measurements can be obtained by solving the difference between the normalized intensity on the front and back defocus planes [12]. The advantage of CS is that the optical path is simple, no decoupling operation is required, the reconstruction of a phase is rapid, and the measuring accuracy for low spatial frequency aberration is high. Moreover, it can detect both the piston error and the tip/tilt error; however, it has the disadvantage of having a small capture range ($\pm\lambda/8$) and being susceptible to atmospheric influences.

At present, only PR, PD, and CS do not introduce non-common path errors in co-phasing detection technologies. Comparing the methods mentioned above, we choose CS as our co-phasing detection method, because it has no non-common optical path error, no decoupling operation, and rapid phase reconstruction. However, CS, similar to PR and PD, has the problem of 2π ambiguity when using monochromatic light measurements. Therefore, this study uses a multi-wavelength measurement method to eliminate the influence of 2π ambiguities and improve the measurement range of co-phasing detection. It is assumed that the tip/tilt error of each segment has been completely removed, and the piston value of each segment satisfies the uniform distribution within the limits of $\pm 11\lambda$, where λ is the operating wavelength.

The residual piston error on the segmented primary mirror will introduce diffraction effects at the in-focus plane and out-of-focus plane. Consequently, the physical optics-based CS method can be applied to measure the piston error of the segments. However, the application of the CS method on ground-based telescopes is limited by atmospheric seeing [13].

Aiming to address the problem that the CS method is easily influenced by atmospheric seeing, this study proposes the method of particle swarm optimization optimized wavelet support vector machine (PSO-WSVM) to solve the influence of atmospheric seeing, and to improve the application scope of the CS method.

To simulate the effects of atmospheric seeing, the images obtained at the CCD are filtered by the long exposure optical transfer function (OTF) of the atmosphere, developed by Fried [14]. The Fried constant selected in this step satisfies a uniform distribution on the interval [0.1, 0.2] (unit: m) [15]. Thus, a casual atmospheric seeing is added to each image. The method proposed in this study should be able to accurately guess the piston error range of each pair of adjacent segments without knowing the Fried constant.

The basic idea of machine learning (ML) is to utilize a different algorithm to generate models from input data on a computer. The concept has evolved to be employed in many different applications, such as expert systems, data mining, image recognition, fault diagnosis, natural language processing, robotics, and biomedical applications [16]–[20]. In recent years, artificial intelligence methods such as SVM, convolutional neural networks (CNNs), self-encoding networks, and anti-neural networks have emerged, some of which have been used for wavefront sensing and have achieved excellent performance.

Dailos *et al.* [21] employed CNNs to predict the piston step values between segments with high accuracy, as well as a broad capture range at visible wavelengths. Li *et al.* [22] proposed a novel approach to construct the feature vector by the in-focal and defocused images, which can get rid of the dependence of the CNN dataset on the imaging target. Nishizaki *et al.* [23] presented a new class of wavefront sensors by extending their design space based on CNN, which simplifies both the optical hardware and image processing in

wavefront sensing. Xu *et al.* [24] proposed a control algorithm based on a deep learning control model to compensate for wavefront aberrations, eliminating the dependence on the deformable mirror response matrix. Paine and Fienup [25] used machine learning operating on a point-spread function to determine a reasonable initial estimate of the wavefront.

In the 1990s, SVM for machine learning was proposed. Since SVM has a principle based on structural risk minimization (SRM), it has unique advantages in working on nonlinear problems with fewer samples and high dimensions [26]. Compared with CNN, the unique advantage of SVM is that the model is small and does not require a heavy computational load. In addition, no special hardware is required in SVM. Complete in theory and straightforward in application, it has gradually become a popular machine learning algorithm that is widely used in pattern recognition, object identification, cancer recognition, image semantic segmentation, automatic text classification and biological sciences [27]–[29].

II. THE PROPOSED MODEL

In this section, we present this novel model in detail. First, we provide CS based on physical optics and introduce atmospheric seeing into the model. Then, we propose an SVM classifier based on the wavelet kernel function. After that, we present the PSO algorithm with the exponential decreasing inertia weight (EDIW) to optimize the parameters of the WSVM. Finally, the flow of the entire algorithm is given in detail.

A. WAVEFRONT SENSING MODEL CONSIDERING ATMOSPHERIC SEEING

Located on the Mauna Kea volcano summit in Hawaii, the Keck telescope is one of the most successful segmented telescopes in the world. The primary mirror of the Keck Telescope is 10 m in diameter and is a mosaic composed of 36 hexagonal segments with a diameter of 1.8 m each. The optical structure of the Keck telescope is used as the optical model of this study. The segmented primary mirror is shown in Figure 4, and the generalized pupil function of it is given by:

$$f(\boldsymbol{\rho}) = \sum_{i=1}^{n} h\left(\boldsymbol{\rho} - \boldsymbol{\rho}_i\right) \exp\left[i\phi\left(\boldsymbol{\rho} - \boldsymbol{\rho}_i\right)\right] \tag{1}$$

where, $h\left(\boldsymbol{\rho} - \boldsymbol{\rho}_i\right)$ is the transmission intensity function of the ith segment and can be given by:

$$h\left(\boldsymbol{\rho} - \boldsymbol{\rho}_i\right) = \begin{cases} 1, & \left|\boldsymbol{\rho} - \boldsymbol{\rho}_i\right| \le a \\ 0, & \left|\boldsymbol{\rho} - \boldsymbol{\rho}_i\right| > a \end{cases} \tag{2}$$

$\boldsymbol{\rho}$ and $\boldsymbol{\rho}_i$ are the coordinate vector on the aperture plane and the central coordinate of the ith segment, respectively. $\phi\left(\boldsymbol{\rho} - \boldsymbol{\rho}_i\right)$ is the phase function of the ith segmented mirror.

Suppose two adjacent mirror segments are located on both sides of the y-axis and the gap between them is arbitrarily small. The phase is ϕ_1 in the field of $x < 0$ while the phase is ϕ_2 in the field of $x > 0$. Thus, the phase difference between the adjacent segments is $\Delta\phi$, where $\Delta\phi = \phi_1 - \phi_2$. The piston error between the adjacent segments will affect the normalized intensity distribution at a distance z from the segmented primary mirror.

The illuminance on the detector plane parallel to the screen with z apart from the screen is the function of x which is given by [30]:

$$\begin{aligned} I(x)/I_0 &= 1 + A\left[\sin(\Delta\phi)\right] + B\left[1 - \cos(\Delta\phi)\right] \\ A &= C(x/\rho) - S(x/\rho) \\ B &= C^2(x/\rho) - S^2(x/\rho) - 1/2 \\ \rho &= \sqrt{\lambda z/2} \end{aligned} \tag{3}$$

where C and S are the Fresnel cosine and sine integrals, $C(x) = \int_0^x \cos\left(\pi t^2/2\right)dt$ and $S(x) = \int_0^x \sin\left(\pi t^2/2\right)dt$ respectively, and ρ appoints the signal width.

Although Equation (3) gives the phase sensing of two adjacent segments with infinite length, it is impossible to simulate the defocused image of a real segmented primary mirror when the piston error and seeing are considered. Therefore, a new approach needs to be explored to achieve our goals. In this study, the Fresnel-Kirchhoff diffraction formula is applied to achieve this goal. This method takes advantage of the idea of convolution to obtain the complex amplitude at a plane where the pupil has propagated a distance z[30]. It is expressed as:

$$U(\boldsymbol{x}, z) \propto FT^{-1}\left\{FT\left[f(\boldsymbol{\rho})\right] e^{-i\boldsymbol{v}^2 z/2k}\right\} \tag{4}$$

where $\boldsymbol{x}$ is a two-dimensional position vector in the detector plane, $\boldsymbol{v}$ is a spatial frequency vector, $U(\boldsymbol{x}, z)$ is the complex amplitude fields at a distance z from the pupil, $f(\boldsymbol{\rho})$ is the complex pupil function and $\boldsymbol{z}$ is the effective propagation distance defined by:

$$\boldsymbol{z} = \frac{(f - l)f}{l} \tag{5}$$

where l is the distance from the focal plane, and f is the focal length of the segmented primary mirror telescope.

The spatial distribution of image intensity on the defocus plane is the square of the complex amplitude distribution modulus.

$$I(\boldsymbol{x}, z) = |U(\boldsymbol{x}, z)|^2 \tag{6}$$

The defocused image obtained on the out-of-focus plane can be obtained by convolving the object in the spatial domain with the PSF of the defocusing optical system.

$$i_{defocus} = o(\boldsymbol{x}) * PFS_{defocus}(\boldsymbol{x}) \tag{7}$$

The above correlation can be converted to the frequency domain which is expressed as follows:

$$I_{defocus} = O(\boldsymbol{x}) * OTF_{defocus}(\boldsymbol{x}) \tag{8}$$

$$\begin{aligned} OTF_{defocus}(\boldsymbol{x}) &= \left|OTF_{defocus}(\boldsymbol{x})\right| e^{iATF_{defocus}(\boldsymbol{x})} \\ &= MTF_{defocus}(\boldsymbol{x})\, e^{iATF_{defocus}(\boldsymbol{x})} \end{aligned} \tag{9}$$

Considering that a ground-based telescope is affected by atmospheric disturbances, the intensity image acquired on the

out-of-plane is filtered by the long exposure optical transfer function (OTF) of the atmosphere, proposed by Fried [31]. The Fried constant randomly selected for the OTF in this study satisfies the principle of uniform random distribution between 0.1 m and 0.2 m statistically. Accordingly, each simulated out-of-focus image is added to the effects of a different atmospheric turbulence. A simulated intensity image acquired with the described method is shown in Figure 4.

The mathematical model of long-exposure MTF based on the Kolmogorov atmospheric turbulence model is chosen in this study to simulate the imaging system MTF [32].

$$\langle MTF_I(\hat{q})\rangle = MTF_I^d(\hat{q}) \exp\left[-3.44\left(\frac{D}{r_0}\hat{q}/2\right)^{5/3}\right] \quad (10)$$

where $\hat{q} = q_{obj}/q_{cut}$, $q_{obj} = 2\pi f, f$ is the spatial frequency, $q_{cut} = kD/(2L_i)$, k is the wavenumber, D is the diameter of the imaging system with a circular aperture, and L_i is the imaging distance.

B. WAVELET SUPPORT VECTOR MACHINE (WSVM)

In 1995, the support vector machine (SVM) was formally proposed by Cortes and Vapnik [25]. SVM is a new machine-learning algorithm, developed based on SRM principles and statistical learning theory. SVM is essentially a form of supervised learning algorithm, which can eliminate the unfavorable factors such as slow convergence, overfitting, and the easy to fall into local minimum value trap of the artificial neural network. Due to its unique advantages such as its small sample size, strong generalization ability, fast learning efficiency, and relatively simple model construction, it has a wide range of applications in classification, feature recognition, and regression prediction.

The basic idea of SVM is to find an optimal separating hyperplane with the maximum margin in the feature space constructed by mapping the input vectors to a higher dimension through some nonlinear mapping and classifying the different type of sample. The hyperplane divided by this principle has the best tolerance to local disturbances, the resulting classification results are the most robust, and the generalization ability to classify the unexemplified samples is the strongest. Based on this idea, the classification accuracy can be improved.

After continuous research and innovation by scholars in recent years, different versions of SVM has evolved, such as the least squares SVM (LS-SVM), the Bayesian SVM, and the fuzzy SVM.

Compared with traditional SVM, LS-SVM adopts an equality constraint rather than the inequality constraint in traditional SVM, which reduces the complexity of the algorithm. LS-SVM converts the quadratic problem into solving linear equations by treating the error squared loss function as the empirical loss of the training set. It increases the convergence speed and computational efficiency for solving a specific problem in practice. Thus, it achieves a better compromise between accuracy and generalization performance.

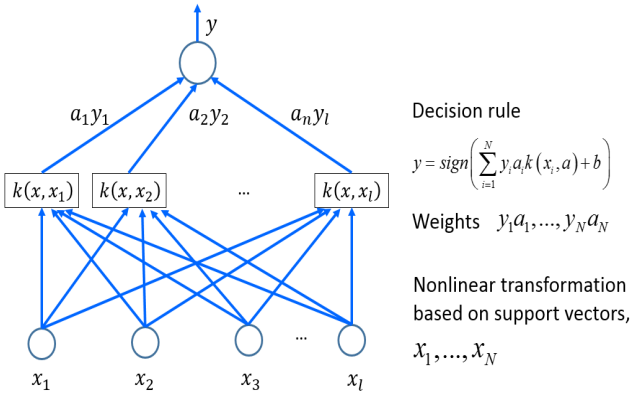

FIGURE 1. The architecture of SVM.

The core idea of machine learning is to estimate the posterior probability of the model, trained as accurately as possible using a limited training set. Given this rule and based on Bayesian theory, Bayesian SVM aims at maximizing the posterior probability of parameter distribution to obtain the optimal parameter value during optimization. Consequently, the selection of the parameter in the model meets the requirements of an objective principle, and the prediction accuracy of SVM is improved to a certain extent.

In some application scenarios, some meaningless data points are generated due to noise, but these data points are not needed for the machine. However, traditional SVM cannot identify and discard these bad points. Given this, a fuzzy membership value is allocated to each input point to improve the noise immunity of the traditional SVM, thus, to reshape the SVM into fuzzy SVM. This method can effectively suppress the adverse effects of noise and outliers on input samples and improve the performance of SVM.

LibSVM is integrated software for support vector classification, (including C-SVC and nu-SVC where C and nu are parameters), regression (epsilon-SVR, nu-SVR) and distribution estimation (one-class SVM) [33].

In the actual application process of SVM, related research shows that the choices of parameters C and the kernel parameter g are closely related to the performance of the algorithm. The key to fully exploit the performance of SVM is to set C and g appropriately.

The choices of penalty factor, kernel function, and kernel parameters are directly related to the performance of the SVM. Solving severe nonlinear problems with SVM do not yield satisfactory results because the traditional kernel function cannot generate a complete base in L_2 space after transform processing.

Suppose there is a set labelled training data $\{(\boldsymbol{x_i}, y_i)\}$ $(i = 1, 2, \ldots, l)$, $\boldsymbol{x_i} \in R^n$ representing the input column vector of the ith training point, and $y_i \in R$ is the output. l represents the number of points in the training set, as shown in Figure 1.

The regression of a series of nonlinear data points $\boldsymbol{x_i} = \left[x_i^1, x_i^2, \ldots, x_i^n\right]$ will become a linear regression problem

after being mapped to a high-dimensional space by the nonlinear mapping function $\varphi(\cdot)$.

$$y = f(\boldsymbol{x}) = \boldsymbol{w} \cdot \varphi(\boldsymbol{x}) + b \tag{11}$$

where, $\boldsymbol{x}$ denotes the SVM input vector, $\boldsymbol{x} = (x_1, x_2, \ldots, x_N)$, y denotes the corresponding SVM output, w represents the connection weight vector of SVM, and b is a constant.

According to the SRM principle, w needs to be selected. The ε-insensitive loss function developed by Vapnik is adopted, which is given by:

$$L_\varepsilon(y - f(x)) = \begin{cases} 0 & |y - f(x)| \le \varepsilon \\ |y - f(x)| - \varepsilon & otherwise \end{cases} \tag{12}$$

where ε is a constant called the tube size and the value of it will affect the number of support vectors.

If the function f can estimate all (x_i, y_i) with precision ε, then searching and finding the smallest w allows the problem to be transformed into a convex optimization problem as follows:

$$\begin{cases} \min \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{N}\left(\xi_i + \xi_i^*\right) \\ s.t. \begin{cases} y_i - \sum_{i=1}^{N} w_i K\left(x_j, x_i\right) - b \le \varepsilon + \xi_i, & j = 1, \ldots, N \\ \sum_{i=1}^{N} w_i K\left(x_j, x_i\right) + b - y_i \le \varepsilon + \xi_i^*, & j = 1, \ldots, N \\ \xi_i, \xi_i^* \ge 0, & j = 1, \ldots, N \end{cases} \end{cases} \tag{13}$$

where ξ_i, ξ_i^* are relaxation variables, b is a constant, $K(\cdot)$ is a nonlinear mapping function, and C is a penalty coefficient. A larger value of C means that a larger penalty will be imposed on the sample point where the training error is higher than ε.

The solution of the dual convex optimization problem described above uses the following Lagrangian function:

$$\begin{aligned} &L\left(\boldsymbol{w}, b, \xi_i, \xi_i^*, \alpha_i, \alpha_i^*, \eta_i, \eta_i^*\right) \\ &= \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{l}\left(\xi_i + \xi_i^*\right) \\ &\quad - \sum_{i=1}^{l}\alpha_i\left(\xi_i + \varepsilon - y_i + \boldsymbol{w}\varphi(x_i) + b\right) \\ &\quad - \sum_{i=1}^{l}\alpha_i^*\left(\xi_i + \varepsilon - y_i + \boldsymbol{w}\varphi(x_i) + b\right) \\ &\quad - \sum_{i=1}^{l}\left(\eta_i\xi_i + \eta_i^*\xi_i^*\right) \end{aligned} \tag{14}$$

where $\alpha_i, \alpha_i^*, \eta_i$ and η_i^* are nonnegative constants and called Lagrangian multipliers.

The objective function $L(\cdot)$ maximizes $\alpha_i, \alpha_i^*, \eta_i, \eta_i^*$, and minimizes $\boldsymbol{w}, b, \xi_i, \xi_i^*$. According to the Karush-Kuhn-Tucker condition, the condition that $L(\cdot)$ takes the extreme value satisfies the following constraints:

$$\begin{aligned} \frac{\partial L}{\partial b} &= \sum_{i=1}^{l}\left(\alpha_i - \alpha_i^*\right) = 0, 0 \le \alpha_i, \alpha_i^* \le C \\ \frac{\partial L}{\partial w} &= w - \sum_{i=1}^{l}\left(\alpha_i - \alpha_i^*\right)\varphi(x_i) = 0 \Rightarrow w \\ &= \sum_{i=1}^{l}\left(\alpha_i - \alpha_i^*\right)\varphi(x_i) \\ \frac{\partial L}{\partial \xi_i^*} &= C - \alpha_i^* - \eta_i^* = 0 \Rightarrow C = \alpha_i^* + \eta_i^* \end{aligned} \tag{15}$$

The above optimization problem is equivalent to solving its dual form as follows:

$$\begin{aligned} \max_{\alpha_i, \alpha_i^*} -\frac{1}{2}\sum_{i,j=1}^{l}\left(\alpha_i - \alpha_i^*\right)\left(\alpha_j - \alpha_j^*\right)\left(\varphi(x_i)\cdot\varphi\left(x_j\right)\right) \\ + \sum_{i=1}^{l}\left(\alpha_i - \alpha_i^*\right)y_i - \sum_{i=1}^{l}\left(\alpha_i + \alpha_i^*\right)\varepsilon \end{aligned} \tag{16}$$

It obeys the constraints:

$$w = \sum_{i}^{l}\left(\alpha_i - \alpha_i^*\right)x_i, 0 \le \alpha_i, \alpha_i^* \le C, i = 1, \ldots, l \tag{17}$$

As a result, the expression function used by the SVM as the regression prediction can be expressed as follows:

$$\begin{aligned} f(x) &= \sum_{i=1}^{l}\left(\alpha_i - \alpha_i^*\right)\left(\varphi(x_i)\varphi(x)\right) + b \\ &= \sum_{i=1}^{l}\left(\alpha_i - \alpha_i^*\right)\kappa(x, x_i) + b \end{aligned} \tag{18}$$

where $\kappa(x, x_i)$ represents the kernel function

The basic idea of SVM to solve nonlinear regression problems is to map the nonlinear regression problems into a high-dimensional space. Thus, the nonlinear problems are transformed into linear problems, and this mapping process is achieved by selecting an appropriate kernel function.

Based on Mercer's condition, every positive definite function $K(x, x_i)$ can be expressed as an inner product $K(x, x_i) = \varphi(x)^T\varphi(x_i)$. Therefore, by choosing a positive definite kernel $\kappa(x, x_i)$ and reformulating the problem in terms of inner products of φ, one can implicitly work in very high dimensional spaces without having to formulate an explicit parametrization of the basis. A popular choice for the kernel function is the Gaussian radial basis function (RBF) kernel. The essential idea of RBF is that the backpropagation learning algorithm applies recursive techniques, which are called random approximations in statistics.

$$\kappa(\boldsymbol{x}, \boldsymbol{x}_i) = \exp\left(-\|\boldsymbol{x} - \boldsymbol{x}_i\|_2^2 / 2\sigma^2\right) \tag{19}$$

where σ ($\sigma > 0$) denotes the bandwidth parameter of the kernel function, and it controls the range of action of the function in the radial direction.

RBF is widely used in regression analysis and pattern recognition and has excellent mapping capabilities. The RBF kernel is associated with an infinite-dimensional feature map. However, the performance is often not ideal when it is applied for approximating complex functions.

Wavelet theory has been widely used in various fields since Morlet proposed it due to its good localization. The modeling of many practical problems can be attributed to the mathematical problem of nonlinear function approximation. The wavelet basis function is a series obtained by transforming and scaling the wavelet decomposition, which has the property of function approximation by wavelet decomposition. Therefore, we can obtain a set of complete bases in L_2 space by expansion and translation of the wavelet basis function and then use it to approximate any nonlinear function. Because it introduces scaling and translation factors, it has more degrees of freedom than the general wavelet decomposition so that it can describe the characteristics of complex functions more accurately. In this study, we applied the wavelet basis function as the kernel function aimed at improving the approximate accuracy of the SVM. The Morlet wavelet basis function is taken as the kernel function and is given by equation (20) [34]:

$$\kappa(x, x_i) = \cos\left(1.75 \times \frac{x - x_i}{\sigma'}\right) \exp\left(\frac{-\|\boldsymbol{x}, \boldsymbol{x}_i\|^2}{2\sigma'^2}\right) \quad (20)$$

where σ' $(\sigma' > 0)$ is the standard deviation of the Morlet wavelet kernel function (MWKF), and it represents the width of the MWKF.

The selection of the input vector and associated parameters of the SVM is critical to the predictive performance of the model being trained. The core SVM hyper-parameters are the penalty parameter C, the tube size ε defined in the epsilon-insensitive function, and the bandwidth parameter σ of the kernel function if the RBF is taken.

The penalty parameter C determines the allowable deviation of the tolerance greater than the threshold ε. When C approaches infinity, the existence of classification error samples is not allowed, and then it becomes a hard margin SVM problem, which leads to overfitting. However, when C tends to 0, it means that we do not consider whether the classification is correct, but that the larger the interval, the better. In this case, a meaningful solution will not be obtained, and the algorithm will also not converge, which results in an underfitting problem. C is a parameter attached to the slack variable introduced by the SVM in order to solve the linear indivisibility of the data set. The slack variable can be understood as a regularization term, and C is the coefficient of the regularization term, which is used to balance empirical risk and model complexity.

The tube size ε ensures the sparsity of dual variables and the existence of a global minimum solution in SVM. The selection of tube size ε is closely related to the generalization capability of the model. The parameter bandwidth σ of the kernel function controls the range of action of the function in the radial direction, and the performance in machine learning is the ability to learn.

The selection of the RBF directly affects the performance of the SVM classifier. Regarding the selection of hyper-parameters for SVMs, previous scholars have done a lot of outstanding research work. Some common strategies that are applied to determine the appropriate hyper-parameters are grid search, K-fold cross-validation (CV), heuristic search, Nelder–Mead search, random search, genetic algorithms (GAs), the PSO algorithm, and pattern search. Currently, the K-fold CV method is widely used in the selection of kernel functions.

For this study, we chose the improved PSO algorithm as the optimization algorithm to address the above problems. The PSO algorithm searches for the best C, σ, and ε parameters by comparing the prediction error in every epoch. The search space of the PSO algorithm forms a three dimensional tensor, one for each parameter.

C. PARTICLE SWARM OPTIMIZATION ALGORITHM FOR WSVM

Inspired by the social flocking behavior of birds, the PSO algorithm is a global optimization algorithm first proposed by Kenney and Eberhart in 1995. The PSO algorithm simulates the clustering and migration behavior of birds during predation. Each particle in the algorithm has individual characteristics, but helps each other to reflect the characteristics of the group [35].

The advantage of the PSO algorithm over other algorithms is that PSO has a memory function, and the knowledge of a good solution is preserved. For example, in GA, the previous knowledge is changed with the change of population, and PSO has no crossover and mutation operation compared with GA. Particles are only updated by internal speed, the principle is more straightforward, there are fewer parameters, and it is easier to implement.

The basic idea of the PSO algorithm is to use the position coordinates and velocity of the current best position and the global value to update the position of the current best position and find a new position. Then it will calculate the corresponding fitness of the new position to evaluate the quality of the solution. The process is shown in Figure 2.

Assume that in an n-dimensional solution space, there is a community of N particles where the ith particle is represented as a one-dimensional vector denoted by $x_i = (x_{i1}, x_{i2}, \ldots, x_{in})$. Suppose the optimal position the ith particle has searched so far is $p_{best_i} = (p_{i1}, p_{i2}, \ldots, p_{in})$. The optimal position searched by the entire particle swarm so far is the global optimal position denoted by $g_{best} = (g_1, g_2, \ldots, g_n)$ and $v_i^k = \left(v_{i1}^k, v_{i2}^k, \ldots, v_{in}^k\right)$ representing the moving rate of the ith particles after k iterations. When these two optimum values are found, the particle swarm

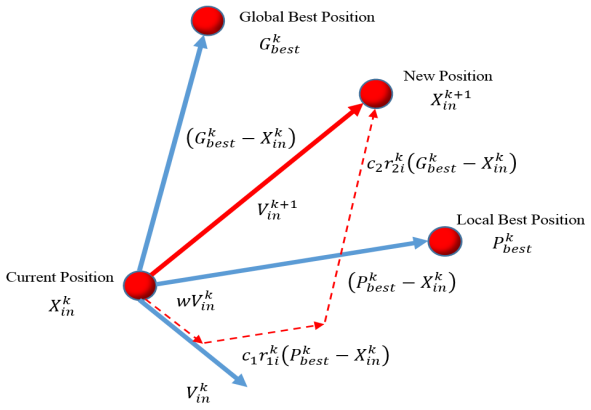

FIGURE 2. Simulated movement of a particle in the search space by updating the position and velocity.

updates its position and velocity according to the following formula:

$$\begin{aligned} v_{id}^k &= w' v_{id}^{k-1} + c_1 rand_1 \left(pbest_{id}^{k-1} - x_{id}^{k-1} \right) \\ &\quad + c_2 rand_2 \left(gbest_d^{k-1} - x_{id}^{k-1} \right) \\ x_{id}^k &= x_{id}^{k-1} + v_{id}^k \end{aligned} \tag{21}$$

where k is a positive integer indicating the number of iterations, i ($i = 1, \ldots, m$) denotes the number of a particle in the set of particles consisting of m particles. d ($d = 1, 2, \ldots, n$) denotes the dimension searched in the n-dimensional solution space. w' represents the inertia weight and w' is a monotone decreasing function in the interval 0.9 to 0.4 during the iterative process [36], c_i ($i = 1, 2$) represents a nonnegative learning factor, and $rand_i$ ($i = 1, 2$) is a normalized random number.

The first term denotes the inertia properties, which indicates the tendency of the particles to maintain their previous speed. The second term shows the cognitive properties of the particles, which represents the tendency of the particles to move toward the current global best solution. The third term signifies the social properties, which shows the group's historical experience of collaboration and knowledge sharing, and represents a trend for the positions of the particles to approach the current optimal position of the group or neighborhood.

The optimization problem of the penalty coefficient and the kernel parameter in SVM can be regarded as the optimization problem in a two-dimensional solution space.

K-fold CV is used in machine learning to work on the problem of overfitting and improve the ability to predict data outside of the training set. It divides the raw data into k groups or folds, where each data subset is taken as a validation set, and the remaining k-1 folds are taken as the training set. In this way, k models are obtained, and the k models are evaluated in the verification set, severally. Next the k-fold CV accuracy is obtained by weighting the error obtained in the k models, which is also called the CV mean of square error (MSE, CV-MSE). CV effectively utilizes limited data, and the results of the evaluation are as close as possible to the performance of the model on the test set. Therefore, CV-MSE is taken as the fitness function value for the trained model in this study. The magnitude of the value of CV-MSE is inversely related to the prediction accuracy.

The parameters of PSO are adjusted to balance the global search and local mining ability of the algorithm. The inertia weight is introduced to the velocity term.

In GA, knowledge is shared between the chromosomes, and the entire population evenly converges towards the optimal domain. However, the flow of information is unidirectional in PSO, that is, only g_{best} gives information to other particles. It makes the search of the entire algorithm follow the current solution, which makes it easy for the algorithm to fall into the local extremum. In order to avoid the above situation of falling into the local extremum, the PSO algorithm should be improved.

By adjusting the parameters of PSO, the global and local exploration abilities of the algorithm can be balanced. The inertia weight w' is introduced to the velocity term, and the w' is linearly or nonlinearly adjusted to the iterative process to balance the globality and convergence speed of the search. The inertia weight w' indicates the capability of the particle to retain the original velocity. The larger w' indicates that the global search ability of the particle is stronger, and the local convergence ability is weaker. Conversely, the smaller w' is, the stronger the local convergence ability of the particle, and the weaker the ability to locate the global minimum.

The inertia weight is adjusted according to the iterative process and the flight condition of the particle to balance the globality of the search and the convergence speed of the algorithm. In the initial stage of optimization, the global search capability should be strengthened while local optimization should be paid more attention to in the later stage. Accordingly, the value of w' should gradually decrease as iterations increase. Therefore, it is necessary to adjust the value of w' dynamically.

The traditional PSO algorithm applies a linear decreasing inertia weight (LDIW). However, when the objective function is a non-convex function, if the initial PSO fails to find the global best advantage, the linearity of w' decreases and falls into the local extremum, resulting in premature convergence. In this study, we utilize EDIW. Its basic idea is making w' change slowly in the early stage of iteration to increase the global search ability of the algorithm and making w' fall faster in the later stage to improve the local search ability of the algorithm, thus finding the global optimal solution.

$$w' = w_{end} \left(\frac{w_{start}}{w_{end}} \right)^{(1-k/k_{\max})} \tag{22}$$

where $w_{start} = 0.9$ and $w_{end} = 0.4$ represent initial and termination weights, respectively, k represents the current number of iterations, $k_{\max}$ denotes the maximum number of iterations set.

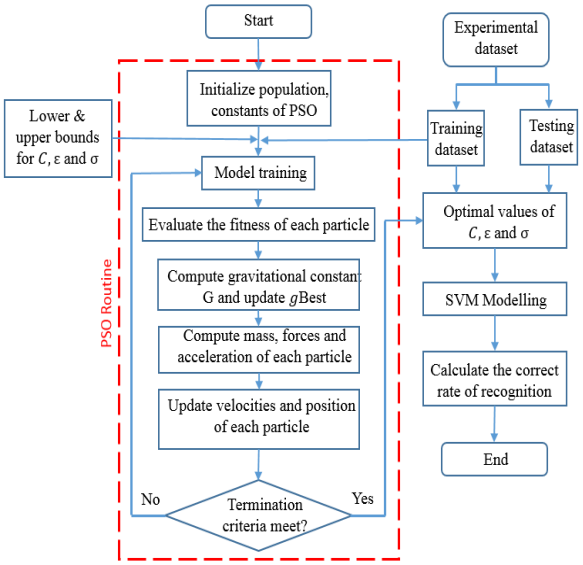

FIGURE 3. Flowchart of EDWI-PSO-WSVM adopted for the study.

The learning factor c_i $(i = 1, 2)$ is often referred to as the acceleration constant to control the moving step size of a particle in an iteration. Generally, $c_1 = c_2 = 2$. The learning factor represents the particle's exploration ability. In the initial stage of optimization, the particles should be given the ability to think and explore more. Therefore, the learning factor should also be adjusted appropriately linearly; c_1 decreases linearly and c_2 increases linearly.

$$c_1 = c_{1s} - \frac{iter\ (c_{1s} - c_{1e})}{iter_{\max}}$$
$$c_2 = c_{2s} + \frac{iter\ (c_{2e} - c_{2s})}{iter_{\max}} \tag{23}$$

where c_{1s} (c_{1e}) and c_{2s} (c_{2e}) denotes initial (or end) values of c_1 and c_2, respectively, *iter* represents the current number of iterations and $iter_{\max}$ denotes the number of iterations at the termination.

In this article, the improved PSO algorithm with EDWI is applied to optimize the critical parameters in the WSVM. The flowchart of the optimization is shown in Figure 3.

III. SIMULATION EXPERIMENTS AND RESULTS

To analyze the performance of our proposed model, we design a simulation experiment and compare the proposed model with some state-of-the-art SVM methods on the same dataset. First, we use the optical system parameters of the Keck telescope to create an optical model of the segmented primary mirror system. Following this, we use this model to get the dataset. Second, we perform statistical processing on the dataset, reducing the input vector dimension of the WSVM and obtaining a new dataset. Finally, we expound the experimental results of the proposed method and compare it with several other state-of-the-art SVM methods.

We aim to control the 2π ambiguity in the wavefront sensing, something may be learned from the idea of redundant measurements, and it is applied to improve the robustness of the algorithm. Given this, we use a set of available lights with the wavelengths chosen from an arithmetic series of numbers. In this study, $\lambda_0 = 700$nm is chosen as the primary wavelength, and three additional shorter wavelengths are also considered: $\lambda_1 = 0.93\lambda_0$, $\lambda_2 = 0.86\lambda_0$, and $\lambda_3 = 0.79\lambda_0$ [20].

TABLE 1. Simulation parameters in the segmented active optics simulation system.

Parameter	Value
Number of segments	36
The focal length of the telescope f	250 m
The defocus length l	4 m
The effective apertures of the primary mirror D	10 m
The edge length of the segment a	0.9 m
Number of pixel lattice of detector	1024×1024
The edge spacing between two adjacent segments d	0.01 m

The piston value of each segment is randomly generated, and the process satisfies the principle of a uniform distribution over the interval $[-11\lambda_0, 11\lambda_0]$. Consequently, the range of the difference between any two adjacent segments is $[-22\lambda_0, 22\lambda_0]$. The parameters selected in the simulated optical system with a segmented primary mirror are shown in Table 1.

Ground-based telescopes are susceptible to atmospheric turbulence, which is the unique property of ground-based telescopes and the reason for the development of adaptive optics. The effect of atmospheric turbulence on the telescope optical system is statistically represented as a change in the optical transfer function (OTF). In the mathematical form, OTF acts as a filter. The perfect parallel wavefront enters the telescope aperture after passing through the atmosphere. Due to piston errors existing on the segmented primary mirror, the wavefront becomes discontinuous. Thus, in the out-of-focal plane, an intensity image is generated. Essentially, the perfect wavefront is filtered twice, once by the atmosphere, and the second time by the effect of the optical system of the considered telescope. In this study, the long exposure OTF of the atmosphere developed by Fried is selected to simulate the filtering effect of atmospheric turbulence on intensity. Here, the Fried parameter is randomly selected, which has a uniform distribution from 0.1 m to 0.2 m. In this way, there is a different atmospheric seeing for each intensity image. A simulated intensity image acquired in a detector is shown in Figure 4.

As shown in Figure 4, the window size is 100 pixels × 100 pixels. The window data captured on the CCD will be further processed to be used as samples for training and testing.

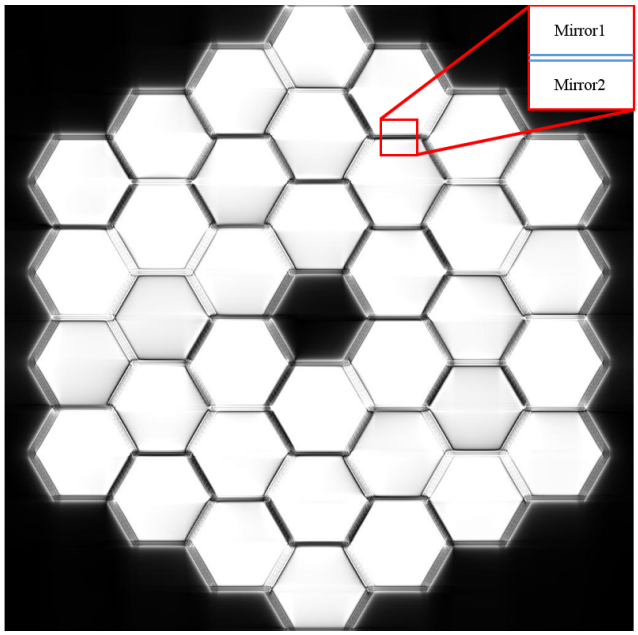

FIGURE 4. An example of the simulated defocus intensity image. A rectangular area centered at the midpoint of the line connecting the geometric center of two adjacent segments is selected as the input data point. The position of this data point is marked with a red solid line window in this figure.

The EDIW-PSO-WSVM proposed in this study works to predict the ambiguity range of the piston error between two adjacent segmented mirrors. Each ambiguity range is assigned a non-repeating integer tag as the training label of the EDIW-PSO-WSVM, and use template matching to obtain the normalized cross-correlation coefficient as input parameters then the machine will perform the task of classification. The interval step of the ambiguity range caused by the different piston errors between each pair of adjacent segments is selected as π. Moreover, the piston errors for each segment satisfy the uniform random distribution on the interval from $-11\lambda_0$ to $11\lambda_0$ and the principle of independent identically distributed hypothesis. The algorithm topology is shown in Figure 5.

First, the piston value of each segment is randomly taken from a uniform distribution over the continuous interval $(-11\lambda_0,\ 11\lambda_0)$, and then the physical optics' method is applied to create a diffraction intensity image collected by the detector at the out-of-focus plane, as shown in Figure 4. The effect of different atmospheric seeing on the acquired intensity image is simulated by introducing a long exposure transfer function and changing the value of the Fried parameter selected randomly from the interval of 0.1 m to 0.2 m. After that, the captured intensity images are cropped into smaller rectangular sample images, and input data is generated by assigning the non-repeating integer labels to the corresponding ambiguity ranges between two adjacent segments. The role of EDIW-PSO-WSVM is to train the machine based on the limited number of these input data points. Once the model is trained, it should be able to predict which labeled range the ambiguity range between each pair of adjacent segmented mirrors falls into. The processing flow for intercepting data is shown in Figure 5.

After the above process, the intensity images obtained by the detector are cropped into smaller rectangular blocks centered on the geometric center of each pair of two adjacent segment edges. An example of the image matting is shown in Figure 4, marked with a solid red line. This matting is to extract the diffraction information between adjacent segments on the out-of-focal plane as the input data of the machine. The data points obtained by the method are a two-dimensional matrix containing the information of diffraction intensity which could only be converted to a one-dimensional vector as an input point for the SVM. It will lead to dimension disaster for SVM if a 3D array is turned into a 2D grayscale image and then converted into a one-dimensional vector. It will increase the computational difficulty of the model. After careful study of the 3D array at the edge of the adjacent segments, it is found that the piston error affects the corrugation shape at the edges and the ridge of the corrugation is parallel to the edge of the segments. Taking this advantage into account, we average the sampling window data along the edge of the mirror, that is, in the direction of the ridge.

Different piston errors are added to the adjacent segments in the model. Following that, a series of waveforms similar to those shown in Figure 5(c) is obtained. In order to suppress the detector noise, the intensity of the pixel point on the detector is averaged along the direction of the gap. The waveform obtained by changing the difference of piston between adjacent two segments ranging from $-\lambda$ to λ with a step size of 0.1λ, as shown in Figure 6. The conclusion can be drawn from Figure 6 that the shape of the waveform changes periodically as the piston difference of the two adjacent segments changes. The variation period of the waveform is an integer multiple of the wavelength of the test light, and the possible difference can be derived from the information of the waveforms.

To reduce the dimension of the EDIW-PSO-WSVM input vectors and improve the recognition accuracy and the efficiency of the model, 22 different ambiguity ranges with four homochromatic lights were simulated with piston error ranging from 0 to 11 $\lambda_i (i = 0, 1, 2, 3)$ and a step size of $0.05\lambda_i (i = 0, 1, 2, 3)$. Then, the normalized processing is performed, and the theoretical data obtained is shown in Figure 7. It can be seen that the signal maps normalized by four monochromatic lights of different wavelength change periodically and the periods of signals corresponding to the different wavelengths are different. It indicates that the one-dimensional vector composed of the signal intensities of four different wavelengths of monochromatic light corresponding to different piston errors is present and unique. Then, the signal strength of four different monochromatic lights can be used as the input vector of WSVM. The piston error information of the segments is fully retained, which significantly improves the calculation efficiency of the model.

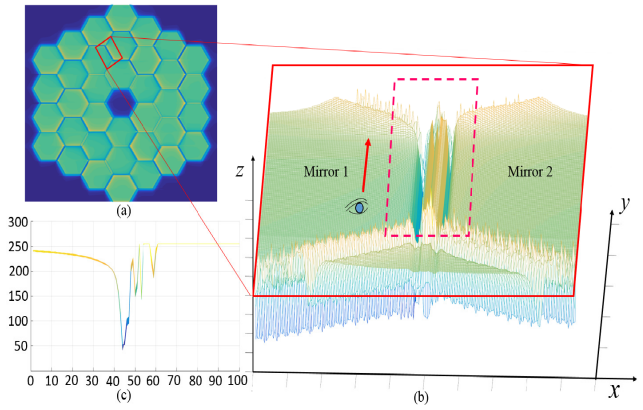

FIGURE 5. Light intensity signal of the image at the edge of the adjacent segments on the CCD at the out-of-focus position: (a) CCD acquisition intensity image at defocused position; (b) 3D map of light intensity information at the cutoff window; (c) The 2D map of the red dotted line in Figure 5(b) is averaged along the direction of the corrugated ridge.

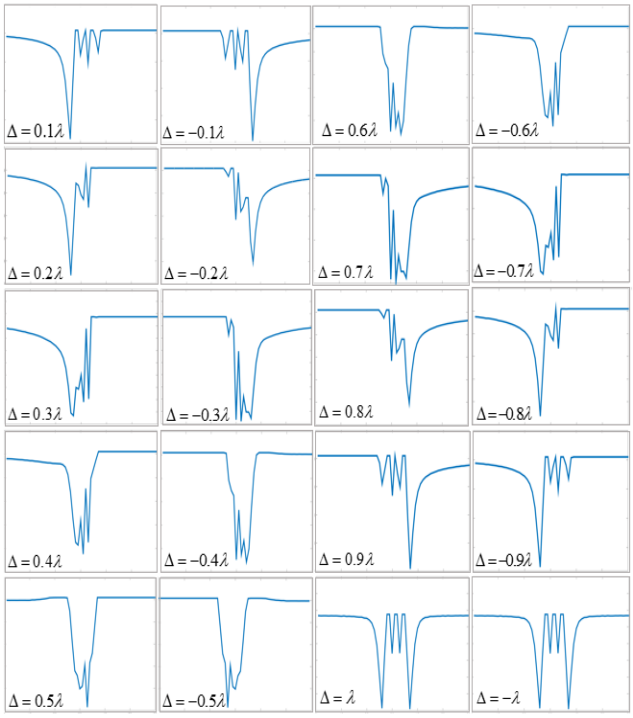

FIGURE 6. The information of the wave corresponding to Figure 5(c) when the two adjacent mirrors have different piston errors.

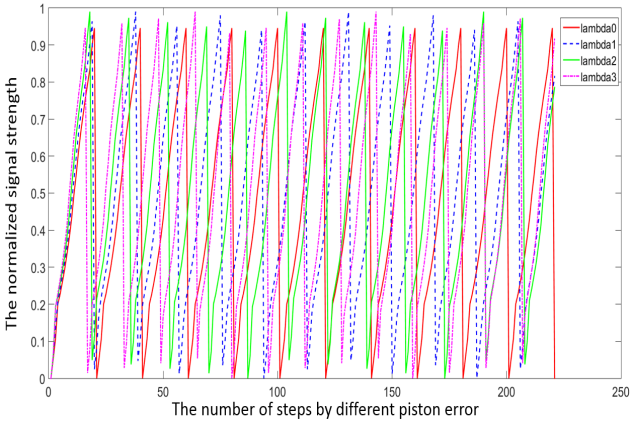

FIGURE 7. The normalized signal strength of four monochromatic lights corresponding to different piston error ranging from 0 to. X-axis and y-axis represent the number of steps (for each step) and the normalized signal strength, respectively.

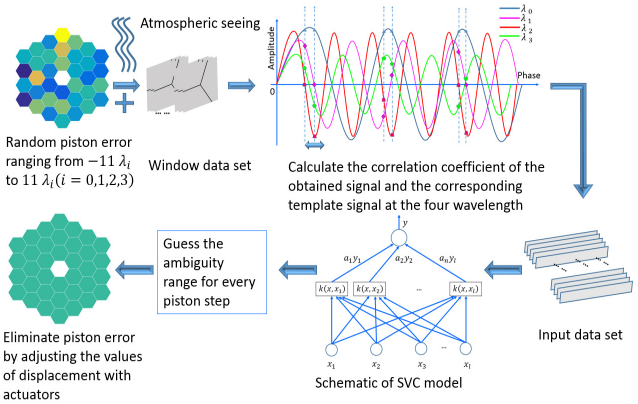

FIGURE 8. Recognition flow of the piston errors of segments.

Given the geometric symmetry of the intensity image generated by the simulation, the data augmentation is used to reduce the computational burden in our simulation process. The sampling window is translated and rotated to acquire more intensity information of the image under diffraction, and the piston error interval of two adjacent segments is recorded as the label to expand the training set.

The flow of the whole simulation experiment is shown in Figure 8. First, a random piston error is added to each segment of the segmented primary mirror system by the MATLAB software, and four different wavelengths of light sources are taken as incident light to obtain their respective light intensities on the detector. Second, the information matrix at the gap of the adjacent segments is extracted, and the matrix is averaged along the gap to obtain the information of the waveform. Third, the normalized cross-correlation coefficient between the waveform is obtained and the template is calculated. The four normalized cross-correlation coefficients calculated for the four different wavelengths of light are input as an input vector in the model of EDIW-PSO-WSVM while the range of the difference in piston between two adjacent segments is taken as the label. A five-fold CV is utilized for training this proposed model, and CV-MSE is chosen as the fitness function value. The improved PSO is used to tune the WSVM hyper-parameters in this model. In this study, data from 8000 samples are used to build up a dataset. From this dataset, the training dataset and testing dataset are generated by random division of the dataset into two parts. In this step, 6000 groups of data are treated as the training dataset while the remaining 2000 groups of data are selected as the testing dataset.

In order to verify the effectiveness of the proposed algorithm, the classification results of this algorithm are compared with other state-of-the-art SVM algorithms. The proposed method is simulated and compared with the

TABLE 2. Classification result.

Methods	RBF			WRBF		
	Time (s)	Train Accuracy (%)	Test Accuracy (%)	Time (s)	Train Accuracy (%)	Test Accuracy (%)
Grid Search	156.0	77.3 (4639/6000)	70.1 (1402/2000)	206.6	78.0 (4681/6000)	72.5 (1450/2000)
PSO	210.6	83.7 (5023/6000)	82.2 (1645/2000)	253.7	89.4 (5365/6000)	89.0 (1779/2000)
GA	1104.4	88.5 (5310/6000)	87.1 (1742/2000)	1206.4	89.3 (5360/6000)	88.8 (1776/2000)
LDIW-PSO	520.5	88.5 (5313/6000)	87.3 (1747/2000)	599.5	89.3 (5359/6000)	88.0 (1759/2000)
BPSO	541.3	92.8 (5568/6000)	90.6 (1813/2000)	618.0	94.4 (5964/6000)	93.2 (1864/2000)
AIW-PSO	859.6	95.8 (5750/6000)	95.3 (1906/2000)	914.8	99.4 (5963/6000)	99.1 (1982/2000)
EDIW-PSO	581.3	99.0 (5941/6000)	98.5 (1970/2000)	611.1	99.6 (5976/6000)	99.5 (1989/2000)

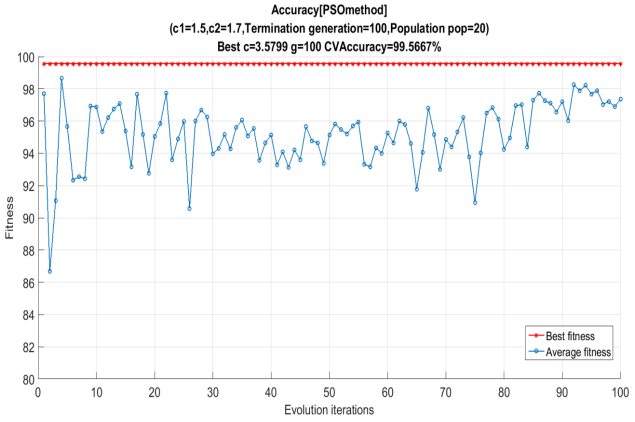

FIGURE 9. Evolution of the best fitness and average fitness in improved PSO method.

following methods: grid search [33], PSO [35], the genetic algorithm (GA) [37], linearly decreasing inertia weight (LDIW-PSO) [38], Butterworth inertia weight strategy (BPSO) [39], adaptive inertia weight algorithm (AIW-PSO) [40].

In order to reduce the accidental error, the average value of 20 simulation experiments is taken. As shown in Table 2, the classification accuracy of the EDIW-PSO-WSVM algorithm is better than that of the other SVM algorithms we compared with. Considering that the proposed algorithm adds computational steps to calculate the kernel matrix in the optimization of parameters c and g, the calculation time is slightly increased, but it is still within the acceptable range.

In this study, the proposed algorithm is applied to the determination of the piston error range between two adjacent segments, which makes up for the 2π ambiguity of the traditional optical detection method and improves on the range of wavefront capture of the curvature wavefront measurement method.

In Figure 9, the classification accuracy of the range of the piston error between adjacent segmented mirrors is shown. As shown in the graph, the proposed EDIW-PSO-WSVM method can quickly find the best parameters in the training model. The average fitness appears to fluctuate in the initial stage of the search, and it gradually converges to about 98 % after nearly 80 evolutionary iterations. The CV accuracy achieved over the training eventually reaches around 99.5 %. The time of training the EDIW-PSO-WSVM to recognize a testing sample is about 0.31 s. The CPU applied in this process is Intel(R) Xeon(R) E5-1620 @ 3.6 GHz. The installed software version of MATLAB is R2017a.

Considering this method is based on image, it is not necessary to build an additional optical path, simultaneously avoiding the introduction of non-common path errors. The recognition accuracy of this method is above 98 %. The training time is about 611 s, and no additional computer configuration is needed. The piston error range of each segment can be recognized in real-time during the observation of a segmented telescope, which solves the 2π ambiguity in traditional optical measurement methods.

IV. CONCLUSION

In this study, we investigated issues related to the piston alignment of segment optical mirrors. We proposed a EDIW-PSO-WSVM algorithm to solve the challenging problems posed by 2π ambiguity and the effects of atmospheric exposure. We considered the following problems caused by traditional methods for wavefront sensing: (1) The traditional wavefront measurement would exhibit a periodic phase variation of 2π using monochromatic light, making it challenging to achieve both high precision and a broad measurement range. (2) Due to the influence of the atmospheric seeing on the ground-based telescope, the traditional optical measurement method will produce an unremovable error, which will affect the measurement accuracy of wavefront sensing. To address these problems, we proposed a EDIW-PSO-WSVM algorithm combining four different wavelengths of the light source. We improved the shortcomings of the traditional wavefront sensing technique, in which it is susceptible to 2π ambiguity. We apply mathematical statistics and image processing techniques to extract features of the image plane as input parameters, and use template matching to obtain the normalized cross-correlation coefficient as input parameters to establish a EDIW-PSO-WSVM classifier. The global search capability and convergence speed are balanced by introducing the EDWI. Thereafter the enhanced PSO algorithm is applied to find the hyper-parameters (c and g) of SVM. The simulation results presented in this paper prove that the proposed EDIW-PSO-WSVM is superior to other state-of-the-art SVM methods. It can distinguish the piston error range between each pair of adjacent segments accurately, and it can exclude the limitation of 2π ambiguity in any interferometry.

Our approach has many advantages over other co-phasing detection techniques. First, our method expands the measurement range of the piston error between adjacent segments to $\pm 11\lambda_0$ and the accuracy of our method is better than $0.05\lambda_0$ RMS, in units of the largest wavelength. Second, our method is adaptable, fast, and does not require iteration

once it is trained. Moreover, our model can be re-trained with subsequent acquisition data to obtain a more accurate model. Finally, our method does not require the provision of an additional optical path, but an imaging detector. Therefore, it can also avoid the introduction of non-common path errors. Furthermore, the simulation experimental results show that the proposed technique is effective for random atmospheric seeing with Fried parameter changes in the range of 0.1 m to 0.2 m. Future work is currently being carried out to verify the adaptability of our method for the tip/tilt errors.

REFERENCES

[1] B. Li, W. Yu, and X. Cui, "Measuring seeing with a Shack–Hartmann wave-front sensor during an active-optics experiment," *Appl. Opt.*, vol. 43, no. 4, pp. 729–734, Feb. 2004.
[2] K. Liu, X. Wang, and Y. Li, "Local slope and curvature tests via wave front modulations in the Shack–Hartmann sensor," *IEEE Photon. Technol. Lett.*, vol. 29, no. 10, pp. 842–845, May 15, 2017.
[3] M. Deprez, C. Bellanger, L. Lombard, B. Watterllier, and J. Primot, "Piston and tilt interferometry for segmented wavefront sensing," *Opt. Lett.*, vol. 41, no. 6, pp. 1078–1081, Mar. 2016.
[4] B. Li, W. Yu, M. Chen, J. Tang, and H. Xian, "Co-phasing experiment of a segmented mirror using a combined broadband and two-wavelength algorithm," *Appl. Opt.*, vol. 56, no. 32, pp. 8871–8879, Nov. 2017.
[5] D. V. Podanchuk, V. P. Dan'ko, A. A. Goloborodko, and N. S. Goloborodko, "Shack–Hartmann wavefront sensor with the precorrected holographic lenslet array," *Optik*, vol. 131, pp. 520–526, Feb. 2017.
[6] S. Esposito, E. Pinna, A. Puglisi, A. Tozzi, and P. Stefanini, "Pyramid sensor for segmented mirror alignment," *Opt. Lett.*, vol. 30, no. 19, pp. 2572–2574, Oct. 2005.
[7] W.-J. Zhou, X. Guan, F. Liu, Y. Yu, H. Zhang, T.-C. Poon, and P. P. Banerjee, "Phase retrieval based on transport of intensity and digital holography," *Appl. Opt.*, vol. 57, no. 1, pp. A229–A234, Jan. 2018.
[8] T. Wu, J. Dong, X. Shao, and S. Gigan, "Imaging through a thin scattering layer and jointly retrieving the point-spread-function using phase-diversity," *Opt. Express*, vol. 25, no. 22, pp. 27182–27194, Oct. 2017.
[9] M. A. van Dam, B. A. Mcleod, and A. H. Bouchez, "Dispersed fringe sensor for the Giant Magellan telescope," *Appl. Opt.*, vol. 55, no. 3, pp. 539–547, Jan. 2016.
[10] J. Jiang and W. Zhao, "Phasing piston error in segmented telescopes," *Opt. Express*, vol. 24, no. 17, pp. 19123–19137, Aug. 2016.
[11] Z. Xie, H. Ma, X. He, X. He, B. Qi, G. Ren, L. Dong, and Y. Tan, "Adaptive piston correction of sparse aperture systems with stochastic parallel gradient descent algorithm," *Opt. Express*, vol. 26, no. 8, pp. 9541–9551, Apr. 2018.
[12] F. Roddier, "Curvature sensing and compensation: A new concept in adaptive optics," *Appl. Opt.*, vol. 27, no. 7, pp. 1223–1225, Apr. 1988.
[13] N. Baba, H. Tomita, and N. Miura, "Phase diversity imaging through atmospheric turbulence," *Opt. Rev.*, vol. 1, no. 2, pp. 308–310, Jan. 1994.
[14] N. Baba and K. Mutoh, "Measurement of telescope aberrations through atmospheric turbulence by use of phase diversity," *Appl. Opt.*, vol. 40, no. 4, pp. 544–552, Feb. 2001.
[15] B. Bordbar, N. H. Farwell, and M. A. Vorontson, "Wavefront phase retrieval with multi-aperture Zernike filter for atmospheric sensing and adaptive optics applications," *Proc. SPIE*, vol. 9982, no. 6, pp. 1–5, Sep. 2016.
[16] Y. Rivenson, T. Liu, Z. Wei, Y. Zhang, K. de Haan, and A. Ozcan, "PhaseStain: The digital staining of label-free quantitative phase microscopy images using deep learning," *Light, Sci. Appl.*, vol. 8, no. 1, pp. 1–11, Feb. 2019.
[17] Q. Zhang, H. Yu, M. Barbiero, B. Wang, and M. Gu, "Artificial neural networks enabled by nanophotonics," *Light, Sci. Appl.*, vol. 8, no. 1, pp. 1–14, May 2019.
[18] B. Rahmani, D. Loterie, G. Konstantinou, D. Psaltis, and C. Moser, "Multimode optical fiber transmission with a deep learning network," *Light, Sci. Appl.*, vol. 7, no. 69, pp. 1–11, Oct. 2018.
[19] Y. Rivenson, Y. B. Zhang, H. Günaydın, D. Teng, and A. Ozcan, "Phase recovery and holographic image reconstruction using deep learning in neural networks," *Light, Sci. Appl.*, vol. 7, no. 141, pp. 1–9, Feb. 2018.
[20] W. Zuo, F. Zhou, Z. Li, and L. Wang, "Multi-resolution CNN and knowledge transfer for candidate classification in lung nodule detection," *IEEE Access*, vol. 7, pp. 32510–32521, 2019.
[21] G. R. Dailos, D. G. Lara, T. S. Juan, and R. R. M. Jose, "Piston alignment of segmented optical mirrors via convolutional neural networks," *Opt. Lett.*, vol. 43, no. 17, pp. 4264–4267, Sep. 2018.
[22] D. Q. Li, S. Y. Xu, D. Wang, and D. J. Yan, "Large-scale piston error detection technology for segmented optical mirrors via convolutional neural networks," *Opt. Lett.*, vol. 44, no. 5, pp. 1170–1173, Mar. 2019.
[23] Y. Nishizaki, M. Valdivia, R. Horisaki, K. Kitaguchi, M. Saito, J. Tanida, and E. Vera, "Deep learning wavefront sensing," *Opt. Express*, vol. 27, no. 1, pp. 240–250, Jan. 2019.
[24] Z. X. Xu, P. Yang, K. Hu, B. Xu, and H. P. Li, "Deep learning control model for adaptive optics systems," *Appl. Opt.*, vol. 58, no. 8, pp. 1998–2008, Mar. 2019.
[25] S. W. Paine and J. R. Fienup, "Machine learning for improved image-based wavefront sensing," *Opt. Lett.*, vol. 43, no. 6, pp. 1235–1238, Mar. 2018.
[26] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Mar. 1995.
[27] J. Liu, Y.-F. Li, and E. Zio, "A SVM framework for fault detection of the braking system in a high speed train," *Mech. Syst. Signal Process.*, vol. 87, pp. 401–409, Mar. 2017.
[28] X. Wang, S. Guan, L. Hua, B. Wang, and X. He, "Classification of spot-welded joint strength using ultrasonic signal time-frequency features and PSO-SVM method," *Ultrasonics*, vol. 91, pp. 161–169, Jan. 2019.
[29] S. Huang, N. Cai, P. P. Pacheco, S. Narrandes, Y. Wang, and W. Xu, "Applications of support vector machine (SVM) learning in cancer genomics," *Cancer Genom. Proteom.*, vol. 15, no. 1, pp. 41–51, 2018.
[30] A. Schumacher and N. Devaney, "Phasing segmented mirrors using defocused images at visible wavelengths," *Monthly Notices Roy. Astronomical Soc.*, vol. 366, no. 2, pp. 537–546, Feb. 2006.
[31] D. L. Fried, "Optical resolution through a randomly inhomogeneous medium for very long and very short exposures," *J. Opt. Soc. Amer.*, vol. 56, no. 10, pp. 1372–1379, Oct. 1966.
[32] Z. J. Yang and M. A. Vorontsov, "Impact of atmospheric turbulence and refractivity on the modulation transfer function of incoherent imaging systems," *Proc. SPIE*, vol. 9846, May 2016, Art. no. 984602.
[33] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machine," *IEEE/ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 4264–4267, Apr. 2011.
[34] J. Yang, J. W. Tian, J. Liu, and F. Wei, "SVM algorithm based on wavelet kernel function for medical image segmentation," *Proc. SPIE*, vol. 7497, Oct. 2009, Art. no. 74971Z.
[35] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. IEEE Int. Conf. Neural Netw.*, Nov./Dec. 2011, vol. 4, no. 8, pp. 1942–1948.
[36] X. Xu, C. Hua, Y. Tang, and X. Guan, "Modeling of the hot metal silicon content in blast furnace using support vector machine optimized by an improved particle swarm optimizer," *Neural Comput. Appl.*, vol. 27, no. 6, pp. 1451–1461, Aug. 2016.
[37] M. Shojafar, S. Javanmardi, S. Abolfazli, and N. Cordeschi, "FUGE: A joint meta-heuristic approach to cloud job scheduling algorithm using fuzzy theory and a genetic method," *Cluster Comput.*, vol. 18, no. 2, pp. 829–844, Jun. 2015.
[38] Y. Shi and R. C. Eberhart, "Empirical study of particle swarm optimization," in *Proc. Congr. Evol. Comput.-(CEC)*, vol. 3, 1999, pp. 1945–1950.
[39] X. Zhu and H. Wang, "A new inertia weight control strategy for particle swarm optimization," *AIP Conf. Proc.*, vol. 1955, Apr. 2018, Art. no. 040095, doi: 10.1063/1.5033759.
[40] X. Wang, G. Wang, and Y. Wu, "An adaptive particle swarm optimization for underwater target tracking in forward looking sonar image sequences," *IEEE Access*, vol. 6, pp. 46833–46843, 2018.

HAIFENG CAO received the B.S. degree from the School of Mechanical Engineering and Automation, Northeastern University, Shenyang, Liaoning, China, in 2015. He is currently pursuing the Ph.D. degree with the University of Chinese Academy of Sciences, Beijing, and the Changchun Institute of Optics, Fine Mechanics, and Physics, Chinese Academy of Sciences, Changchun, Jilin. His current research interests include active optics and machine learning.

JINGXU ZHANG received the B.S. and M.S. degrees in precision mechanism from the Beijing University of Posts and Telecommunications, in 1987 and 1990, respectively, and the Ph.D. degree from the Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, in 2008. He is currently a Research Fellow with the Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences. His main research interest includes general structure design of ground-based large-aperture telescopes.

FEI YANG received the B.S. degree from the Harbin Institute of Technology, Harbin, Heilongjiang, in 2003, the M.S. degree from the Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, Jilin, in 2009, and the Ph.D. degree from the Changchun University of Science and Technology, Changchun, in 2017. He is currently an Associate Professor with the Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences. His main research interest includes optical and mechanical systems of ground-based large aperture telescopes.

QICHANG AN received the B.S. degree from the School of Engineering Science, University of Science and Technology of China, Hefei, Anhui, in 2011, the M.S. degree from the University of Chinese Academy of Sciences, Beijing, in 2014, and Ph.D. degree from the Changchun Institute of Optics, Fine Mechanics, and Physics, Chinese Academy of Sciences, Changchun, Jilin, in 2018. He is currently a Research Assistant with the Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences. His main research interest includes advanced wavefront detection.

YE WANG received the B.S. degree from the School of Materials Science and Engineering, South China University of Technology, Guangzhou, Guangdong, China, in 2015. She is currently pursuing the Ph.D. degree with the University of Chinese Academy of Sciences in Beijing, the Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, Jilin. Her current research interests include photo detector and femto-laser microfabrication.