Research Article
Vol. 57, No. 11 / 10 April 2018 / Applied Optics 2747

applied optics

Efficient solution to the stagnation problem of the particle swarm optimization algorithm for phase diversity

Xin Qi,[1,2] Guohao Ju,[1,*] and Shuyan Xu[1]

[1]Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, Jilin 130033, China
[2]University of Chinese Academy of Sciences, Beijing 100049, China
*Corresponding author: juguohao123@163.com

Received 27 December 2017; revised 5 March 2018; accepted 7 March 2018; posted 8 March 2018 (Doc. ID 318507); published 3 April 2018

The phase diversity (PD) technique needs optimization algorithms to minimize the error metric and find the global minimum. Particle swarm optimization (PSO) is very suitable for PD due to its simple structure, fast convergence, and global searching ability. However, the traditional PSO algorithm for PD still suffers from the stagnation problem (premature convergence), which can result in a wrong solution. In this paper, the stagnation problem of the traditional PSO algorithm for PD is illustrated first. Then, an explicit strategy is proposed to solve this problem, based on an in-depth understanding of the inherent optimization mechanism of the PSO algorithm. Specifically, a criterion is proposed to detect premature convergence; then a redistributing mechanism is proposed to prevent premature convergence. To improve the efficiency of this redistributing mechanism, randomized Halton sequences are further introduced to ensure the uniform distribution and randomness of the redistributed particles in the search space. Simulation results show that this strategy can effectively solve the stagnation problem of the PSO algorithm for PD, especially for large-scale and high-dimension wavefront sensing and noisy conditions. This work is further verified by an experiment. This work can improve the robustness and performance of PD wavefront sensing.

OCIS codes: (010.7350) Wave-front sensing; (100.5070) Phase retrieval; (000.3860) Mathematical methods in physics; (220.1080) Active or adaptive optics.

https://doi.org/10.1364/AO.57.002747

1. INTRODUCTION

Phase diversity (PD) is a well-known image-based technique that can jointly estimate the object and phase aberrations simultaneously [1–3]. Compared to traditional wavefront sensors, PD offers several advantages, such as a low requirement for optical hardware and feasibility of both point source and extended objects. Therefore, since the inception of PD about three decades ago, this technique has been widely used in various domains, such as wavefront sensing in adaptive and active optics [4–7], biological microscopy imaging [8], and quality control of laser beams [9].

One key point of PD is to develop a suitable optimization algorithm to minimize the error metric (objective function). This error metric is built by evaluating the deviation between the actually obtained intensity images (with a known diversity phase between them) and the theoretical intensity images with the theory of Fourier optics. Then the problem of reconstructing aberration coefficients of the wavefront phase is transferred to searching the coefficient set for which the error metric is a global minimum. Many standard gradient-based nonlinear optimization algorithms [10–13] have been used by researchers to complete this work, such as steepest decent (SD) algorithm [10,11], conjugate-gradient (CG) algorithm [5], and quasi-Newton algorithm (BFGS) [12,13], etc. However, the search of the optimal parameter set with these gradient-based optimization algorithms can easily be trapped in a local minimum that is not the true solution. The main reason for this is that the searching direction of them mainly depends on the derivative information of the error metric.

From this perspective, population-based optimization algorithms are more suitable for PD than those gradient-based optimization algorithms [14,15], for they rely directly upon objective function values rather than derivative information. Also, various randomly initialized agents are dispersed throughout the search space, which allows for searching the optimal parameter set in multiple regions of the search space simultaneously. Compared to other population-based optimization algorithms, such as the genetic algorithm (GA) [16,17] and simulated annealing (SA) [18,19], the particle swarm optimization (PSO) algorithm has simple structure, high convergence efficiency, and fast searching ability due to its parallel search mechanism [20–23]. Therefore, the PSO algorithm is very suitable for PD, which is essentially a multi-dimensional complex optimization problem and has a requirement for searching efficiency. Recently, PSO has been introduced to the area of adaptive optics (AO) microscopy [24] and PD [25] as a global optimization algorithm.

However, we should note that the traditional PSO algorithm for PD still suffers from the stagnation problem (premature convergence) [26,27]. This problem will become even more severe with the increase in the noise level and the scale and dimension of the aberration coefficients set to be searched. In these cases, the objective function values of the local optimum solution (a solution that is optimal within a neighboring set of candidate solutions) and global optimum solution (the optimal solution among all possible solutions) can be similar (due to noise) and spatially distant from each other (due to the large scale and high dimension of the search space). If random initial conditions bring the search closer to the local optimum solution, the individual particle optima (the optimal solution for an individual particle) and group optima (the optimal solution for all particles) will converge towards the local optimum solution when using the traditional PSO algorithm and a wrong optimization result will be obtained. While the PSO algorithm has been introduced to the area of PD wavefront sensing in Ref. [25], the emphasis is placed only on the efficiency, and the stagnation problem has been neglected.

If the ability to search the global optimum solution cannot be guaranteed, the fast convergence efficiency of PSO for PD will lose its sense. To solve the stagnation problem of PSO for PD, an explicit strategy is proposed in this paper based on a deep understanding of the inherent optimization mechanism of the PSO algorithm. One basic character of the PSO algorithm is that the particle swarm will gravitate towards the optimum solution (may be a local minimum), instead of randomly looking for the solution individually. With this knowledge, we first propose a criterion to detect the stagnation problem (premature convergence). Then we introduce a redistributing mechanism to liberate the particles from the local minimum (i.e., to prevent premature convergence). To make the redistributed particles cover the search space more efficiently and improve the efficiency of the redistributing mechanism, randomized Halton sequences are further introduced to ensure the uniform distribution and randomness of the initial particles and the redistributed particles in the search space. Halton sequences are a kind of low-discrepancy sequence. The quasi-random numbers generated by low-discrepancy sequences can cover the domain of interest quickly and evenly [28]. Simulation results show that this strategy can effectively prevent premature convergence and solve the stagnation problem of the PSO algorithm for PD, especially for large-scale and high-dimension wavefront sensing and noisy conditions. This work is further verified by an experiment.

This paper is organized as follows. In Section 2, the principle of the PD is reviewed. In Section 3, the traditional PSO algorithm is reviewed, and the stagnation problem of it for PD is illustrated. Then an explicit strategy is proposed to solve the stagnation problem of PSO for PD in Section 4. In Section 5, simulations and the experiment are performed to demonstrate the effectiveness of the introduced strategy. In Section 6, this paper is concluded.

2. REVIEW OF PHASE DIVERSITY TECHNIQUE

In this section, we will review the principle of the PD technique [1–3]. Let us suppose that the object is illuminated with noncoherent quasi-monochromatic light, and the imaging system is a linear shift-invariant system. The intensity distribution of the image plane with Gaussian noise can be modeled as

$$d_k(\mathbf{r}) = o(\mathbf{r}) * s_k(\mathbf{r}) + n_k(\mathbf{r}), \tag{1}$$

where $*$ denotes the convolution operation, $\mathbf{r}$ is a two-dimensional position vector in the image plane, $o(\mathbf{r})$ is the object, $d_k(\mathbf{r})$ is the kth detected diversity image, $s_k(\mathbf{r})$ is the kth point-spread function (PSF), and $n_k(\mathbf{r})$ is the Gaussian noise in the kth image. With the condition of near field, the PSF associated with the kth diversity image is given by

$$s_k(\mathbf{r}) = |F^{-1}\{P(\boldsymbol{\rho})\exp[i\varphi_k(\boldsymbol{\rho})]\}|^2, \tag{2}$$

where F^{-1} denotes the inverse Fourier transform, $\boldsymbol{\rho}$ is a two-dimensional position vector in the pupil plane, $P(\boldsymbol{\rho})$ is the binary aperture function with values of 1 inside the pupil and 0 outside, and $\varphi_k(\boldsymbol{\rho})$ represents the wavefront phase associated with the kth intensity measurements. In the current PD algorithm, $\varphi_k(\boldsymbol{\rho})$ can further be expressed as

$$\varphi_k(\boldsymbol{\rho}) = \phi(\boldsymbol{\rho}) + \Delta_k(\boldsymbol{\rho}), \tag{3}$$

where $\phi(\boldsymbol{\rho})$ is the unknown wavefront aberration to be estimated, and $\Delta_k(\boldsymbol{\rho})$ is the deliberately introduced kth PD, which is usually known to us.

To evaluate the difference between the diversity images predicted by the imaging model of the optical system and those directly collected, an error metric can be defined as

$$E = \sum_{k=1}^{K}\sum_{\mathbf{r}}[d_k(\mathbf{r}) - o(\mathbf{r}) * s_k(\mathbf{r})]^2. \tag{4}$$

According to the convolution theorem and the Parseval theorem, this error metric can be rewritten in the frequency domain as

$$E = \sum_{k=1}^{K}\sum_{\mathbf{u}}[D_k(\mathbf{u}) - O(\mathbf{u}) * S_k(\mathbf{u})]^2, \tag{5}$$

where $\mathbf{u}$ is a two-dimensional spatial frequency coordinate, and $D_k(\mathbf{u})$, $O(\mathbf{u})$, and $S_k(\mathbf{u})$ denote the Fourier transforms of $d_k(\mathbf{r})$, $o(\mathbf{r})$, and $s_k(\mathbf{r})$, respectively.

In order to reduce the dimensions of the parameter space over which a numerical optimization is performed, the partial differential of the error metric E with respect to the object frequency spectrum O is set to zero. In this case, we can obtain

$$O(\mathbf{u}) = \frac{\sum_{k=1}^{K} D_k(\mathbf{u})S_k^*(\mathbf{u})}{\sum_{k=1}^{K}|S_k(\mathbf{u})|^2}, \qquad \sum_{k=1}^{K}|S_k(\mathbf{u})|^2 \neq 0. \tag{6}$$

$S_k^*(\mathbf{u})$ in Eq. (6) represents the complex conjugate of $S_k(\mathbf{u})$.

Substitution of $O(\mathbf{u})$ into E yields

$$E = \sum_{k=1}^{K}\sum_{\mathbf{u}} |D_k(\mathbf{u})|^2 - \sum_{\mathbf{u}\in\chi} \frac{\left|\sum_{k=1}^{K} D_k(\mathbf{u})S_k^*(\mathbf{u})\right|^2}{\sum_{k=1}^{K} |S_k(\mathbf{u})|^2}, \quad (7)$$

where χ represents a set of spatial frequencies $\mathbf{u}$,

$$\chi = \left\{\mathbf{u}\left|\sum_{k=1}^{K} |S_k(\mathbf{u})|^2 \neq 0\right.\right\}. \quad (8)$$

The unknown wavefront aberration is usually expanded on a finite set of fringe Zernike polynomials,

$$\phi(\boldsymbol{\rho}) = \sum_{j=4}^{N} C_j Z_j(\boldsymbol{\rho}). \quad (9)$$

The coefficients C_1–C_3 stand for piston, tip, and tilt of the wavefront aberration, which have no effect on the quality of the image. The error metric E is therefore defined on a multidimensional parameter space:

$$\mathbf{a} = [C_4, C_5, \ldots, C_N]. \quad (10)$$

For a group of given parameters $\mathbf{a}$, the error metric $E(\mathbf{a})$ can be calculated. The problem of reconstructing aberration coefficients of the wavefront phase can be transferred to searching the coefficient set for which the error metric presented by Eq. (7) is a global minimum. In other words, one of the key points in PD is to develop a suitable optimization algorithm to minimize the error metric and find the global optimum solution.

3. STAGNATION PROBLEM OF THE TRADITIONAL PSO ALGORITHM FOR PHASE DIVERSITY

A. Review of the Traditional PSO Algorithm

PSO was first proposed by Kennedy and Eberhart in 1995 [20]. The PSO search procedures are based on the swarm concept, which is a group of individuals that are able to optimize a certain objective function. Each individual can send information to another and ultimately allow the entire group to move towards the same object or in the same direction. It is a way to simulate the behavior of individuals of the species who work for the benefit of the entire group.

The basic model of PSO calculation is shown below. A particle swarm is randomly initialized and dispersed throughout the search space. Each particle logically chooses the method it will move itself by referring to the individual's best experience and the group's best experience. In this paper, the "traditional PSO algorithm" particularly refers to the variant of the PSO algorithm proposed by Clerc [22], for this variant seems more widely applied at present. This variant of PSO mainly includes two parts, which are shown below.

1. Velocity Update

The velocity of each particle is updated by the following equation:

$$v_i^{k+1} = w[v_i^k + c_1 r_1 (P_i^k - s_i^k) + c_2 r_2 (G^k - s_i^k)], \quad (11)$$

where c_1 and c_2 are learning factors of PSO, r_1 and r_2 are uniformly distributed random numbers between 0 and 1, P_i^k and G^k are individual best optima for particle i and group optima after k iterations, respectively, s_i^k represents the position of particle i in iterative k, v_i^k and v_i^{k+1} are velocities of particle i in iterative k and $k+1$, respectively, and w is weighting factor, which can further be expressed as

$$w = \frac{2}{\left|2 - K - \sqrt{K^2 - 4K}\right|}, \quad K = c_1 + c_2, \quad K > 4. \quad (12)$$

2. Position Update

The position of each particle is updated by the following equation:

$$s_i^{k+1} = s_i^k + v_i^{k+1}. \quad (13)$$

After continuous iterations, the particle swarm will gravitate towards the optimum solution (instead of randomly looking for the solution). The optimization process will terminate when the stop criterion is satisfied.

B. Illustration of the Stagnation Problem of the Traditional PSO Algorithm for PD

In this section, we will reveal the problem of the traditional PSO algorithm when it is applied to solving the PD problem, especially in the noisy condition. Specifically, we will show that the accuracy of this algorithm is very sensitive to noise and at a certain noise level, this algorithm can become invalidated and no longer be used for PD.

We can see from Section 3.A that there is not a mechanism for preventing being trapped in a local minimum in the traditional PSO algorithm. The objective function of the local optimum solution and global optimum solution can be similar and spatially distant from each other. In this case, if random initial conditions bring the search closer to the local optimum solution, the individual particle optima and group optima will converge towards the local optimum solution and cannot jump out, therefore losing its ability to search for a global optimum solution. This problem will become even more severe with the increase in the noise level and the scale and dimension of the aberration coefficients set to be searched.

In the following, we will perform some simulations to illustrate the stagnation problem of the traditional PSO algorithm for PD. Here, we consider only the situation of point source. Three cases will be considered in this simulation:

(1) In the first case, we consider five aberration coefficients, i.e., C4–C8 of the fringe Zernike coefficients (corresponding to defocus, astigmatism, and coma aberrations). We randomly introduce 100 sets of coefficients within certain range. For each set of aberration coefficients, we can use them to generate the in-focus and defocus PSF images with Fourier optics. With these two intensity images and the principle of the PD presented in Section 2, we can use the traditional PSO algorithm to reconstruct the five aberration coefficients. In this case, two noise conditions will further be considered. Specifically, a certain intensity of additive zero-mean Gaussian white noise will be added to each PSF, where the noise is specified by the ratio of the standard deviation of the noise to the peak value in the noiseless PSF image [5]. The specific range of the coefficients considered in this case is $[-0.1\lambda, 0.1\lambda](\lambda = 500\ \text{nm})$. The two

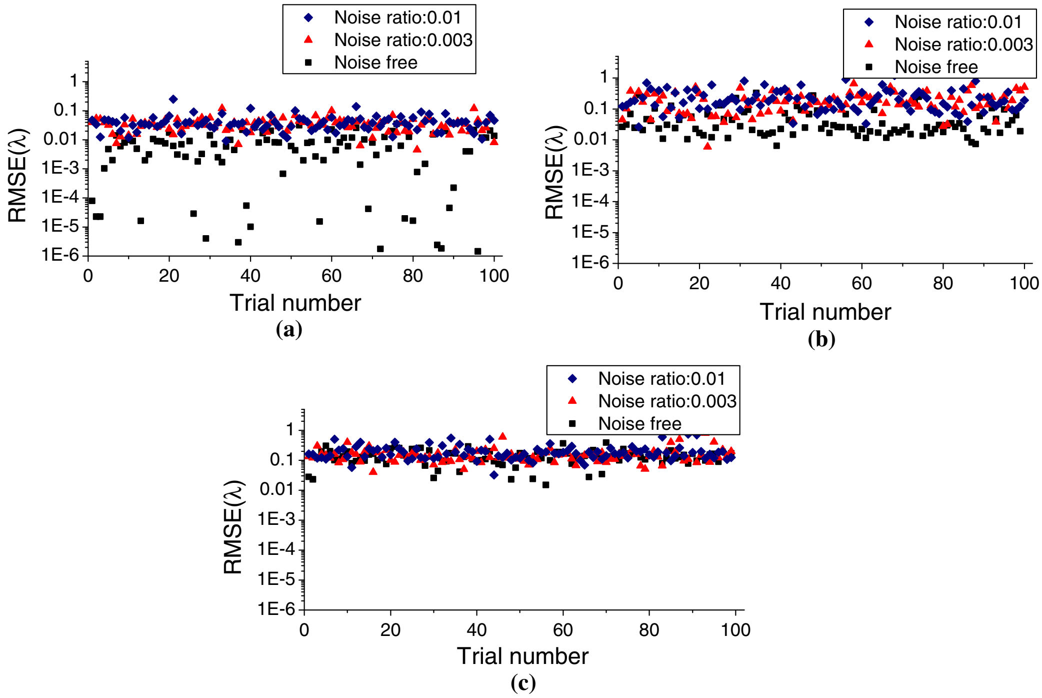

Fig. 1. Illustration of the stagnation problem of traditional PSO for PD under different conditions. In (a), the range of aberration coefficients to be searched is $[-0.1\lambda, 0.1\lambda](\lambda = 500\ \text{nm})$, and the dimension is 5. In (b), the range is increased to $[-0.5\lambda, 0.5\lambda]$, and the dimension is still 5. In (c), the range is $[-0.3\lambda, 0.3\lambda]$, while the dimension is increased to 8. Different noise conditions are also considered in each case. We can recognize that the traditional PSO algorithm for PD is sensitive to noise and the range and dimension of the search space.

noise ratios considered in this case are 0.003 and 0.01, respectively.

(2) The second case considered here is the same as the first case, except that the range of the aberration coefficients is increased to $[-0.5\lambda, 0.5\lambda]$.

(3) The range of each aberration coefficient considered in this case is $[-0.3\lambda, 0.3\lambda]$, a little smaller than the second case. However, the number of coefficients considered is increased to 8, i.e., C4–C11 of the fringe Zernike coefficients (corresponding to defocus, astigmatism, coma, spherical, and trefoil aberrations).

The root mean square error (RMSE) between the real aberration coefficients and those recovered by the traditional PSO algorithm is used to evaluate the accuracy of the method, which is expressed as

$$\text{RMSE} = \sqrt{\frac{\sum_{j=4}^{m+4-1} (C_j^{(1)} - C_j^{(0)})^2}{m}}, \tag{14}$$

where m is the number of the aberration coefficients that are considered, and $C_j^{(1)}$ and $C_j^{(0)}$ are the jth recovered and introduced aberration coefficients, respectively. Here, we do not use the value of the fitness function to evaluate the accuracy, for it cannot give an intuitive impression on the accuracy of the recovered coefficients.

In the traditional PSO algorithm, the learning factors c_1 and c_2 are usually set as $c_1 = c_2 = 2.05$ [21,25]. The population size is 40, and the maximum number of iterations is 400. The results of the three cases are shown in Fig. 1, where the "trial number" represents the index of the simulations. We can draw the following results from Fig. 1:

(1) The traditional PSO algorithm for PD is sensitive to noise. We can see from Fig. 1(a) that in the absence of noise, the accuracy is acceptable, which decreases apparently in the presence of noise. We can also recognize an apparent decrease in accuracy in Fig. 1(b) due to noise. The underlying reason is that in the presence of noise, the distribution of the error metric over the search space becomes more complicated, and the algorithm becomes more likely to be trapped in the local minimum.

(2) The traditional PSO algorithm for PD is sensitive to the range of aberration coefficients. Comparing the noise-free cases of Figs. 1(a) and 1(b) we can see that the accuracy decreases apparently with the increase in the range. One of the underlying reasons is that it becomes harder to find the true global minimum with the increase in the range of the search space, and the algorithm becomes more susceptible to the stagnation problem.

(3) The traditional PSO algorithm for PD is also very sensitive to the dimension of search space. Comparing the noise-free cases of Figs. 1(b) and 1(c), we can see that the accuracy further decreases apparently with the increase in the dimension of the search space.

Also, we may hardly recognize any difference when the noise ratio increases from 0.003 to 0.01. The reason is that the traditional PSO algorithm for PD is very susceptible to noise. In both cases, the traditional PSO algorithm for PD cannot find the true global minimum, and therefore the results are similar.

4. EFFICIENT SOLUTION TO THE STAGNATION PROBLEM OF THE TRADITIONAL PSO ALGORITHM FOR PD

In this section, we will propose an explicit strategy to solve the stagnation problem of the traditional PSO algorithm for PD, based on a deep understanding of the inherent optimization mechanism of the PSO algorithm. Specifically, we introduce a redistributing mechanism to PSO to detect and prevent premature convergence. Also, to complement the redistributing mechanism, randomized Halton sequences are further introduced to ensure the uniform distribution and randomness of the initial particles and the redistributed particles in the search space. Some discussions are also presented to help understand the strategy proposed in this section.

A. Redistributing Mechanism for Detecting and Preventing Premature Convergence

The redistributing mechanism contains two parts of contents, i.e., when to redistribute the particles and how to redistribute the particles. We redistribute some particles when we detect the premature convergence. The goal of this redistributing mechanism is to liberate particles from the state of premature convergence, i.e., to prevent premature convergence.

Here, we first discuss how to detect the premature convergence. According to the inherent optimization mechanism of the PSO algorithm, there exists an information exchange among different particles, and the particle swarm will gravitate towards the optimum solution (may be a local minimum), instead of randomly looking for the solution individually. All particles are pulled on all dimensions toward the optimum solution via update equations, i.e., Eqs. (11) and (13). Therefore, after enough times of iterations, the positions of the particles in the search space will be clustered around the optimum solution. In this case, the standard deviation of the error metric (objective function) corresponding to the particles should be very small. Therefore, we can detect the premature convergence of the PSO algorithm by analyzing the standard deviation of the error metric. Specifically, we can first set a proper threshold. If the standard deviation of the error metric is smaller than this threshold, we can assume that premature convergence has taken place. This stagnation criterion can be expressed as

$$\begin{cases} \text{YES}, & \text{if } \sigma_E < T, \\ \text{NO}, & \text{if } \sigma_E \geq T, \end{cases} \tag{15}$$

where "YES" means premature convergence or stagnation has happened, "NO" denotes that no premature convergence happens, T is a threshold, and σ_E is the standard deviation of the error metrics of the particles, which can be expressed as

$$\sigma_E = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(E_i - \overline{E})^2}, \tag{16}$$

where N is the number of the particles in the search space, E_i is the error metric corresponding to particle i, and $\overline{E}$ is the mean value of the error metrics of all the particles.

Also, we can also detect premature convergence by analyzing the change of the mean value of the error metrics of all the particles $\overline{E}$ between two iterations, for the mean value of the error metrics of all the particles should become stable when premature convergence happens. Therefore, we can have another criterion for detecting premature convergence, which can be expressed as

$$\begin{cases} \text{YES}, & \text{if } \Delta\overline{E} < T_1, \\ \text{NO}, & \text{if } \Delta\overline{E} \geq T_1, \end{cases} \tag{17}$$

where $\Delta\overline{E}$ represents the change of the $\overline{E}$ between two iterations, and T_1 is a threshold. Both of these two criteria can be used to detect premature convergence.

After discussing how to detect premature convergence, we will continue to discuss how to prevent premature convergence. As mentioned before, the traditional PSO algorithm lacks a mechanism to escape from the local minimum. If no particle encounters a better solution over a period of time, the swarm will continually move closer to the unchanged optimum solution (may be a local minimum). Here, we present a redistributing mechanism to liberate particles from the state of premature convergence. Specifically, when we detect premature convergence, we redistribute those particles close to the optimum position in the search place while reserving the current optimum solution.

Here, the Euclidean distance is used to measure the distance between two particles in the search space, which can be expressed as

$$l(\vec{x}_i, \vec{x}_j) = \|\vec{x}_i - \vec{x}_j\|, \tag{18}$$

where the two vectors, $\vec{x}_i$ and $\vec{x}_j$, represent the positions of two particles (particle i and j) in the search space, $l(\vec{x}_i, \vec{x}_j)$ represents the Euclidean distance between these two particles, and $\|\cdot\|$ is the Euclidean norm. The redistributing criterion for each particle can be express as

$$\begin{cases} \text{YES}, & \text{if } 0 < l(\vec{x}_i, \vec{x}_{\text{OP}}) \leq R, \\ \text{NO}, & \text{else}, \end{cases} \tag{19}$$

where the vector $\vec{x}_{\text{OP}}$ represents the position of the current optimum solution, "YES" means that this particle should be redistributed, "NO" denotes that this particle can be reserved, and R is the radius of the adjacent area of the optimum solution. The particles in this area will be randomly redistributed, except the particle representing the current optimum solution. This redistributing mechanism for preventing premature convergence is illustrated in Fig. 2 (we do not show how to detect premature convergence).

One reason that the current optimum solution is reserved is that if this solution is actually the true global optimum, we can guarantee that the particles will definitely converge to this solution again. The reason that those particles outside of the adjacent area of the optimum solution are retained is that these particles are a little far from the optimum solution. Therefore, they still have the possibility of reaching the true global optimum. The values of T and R should be carefully selected. Redistributing too early did not allow for the desired degree of solution refinement, while redistributing too late means wasting time in a stagnated state.

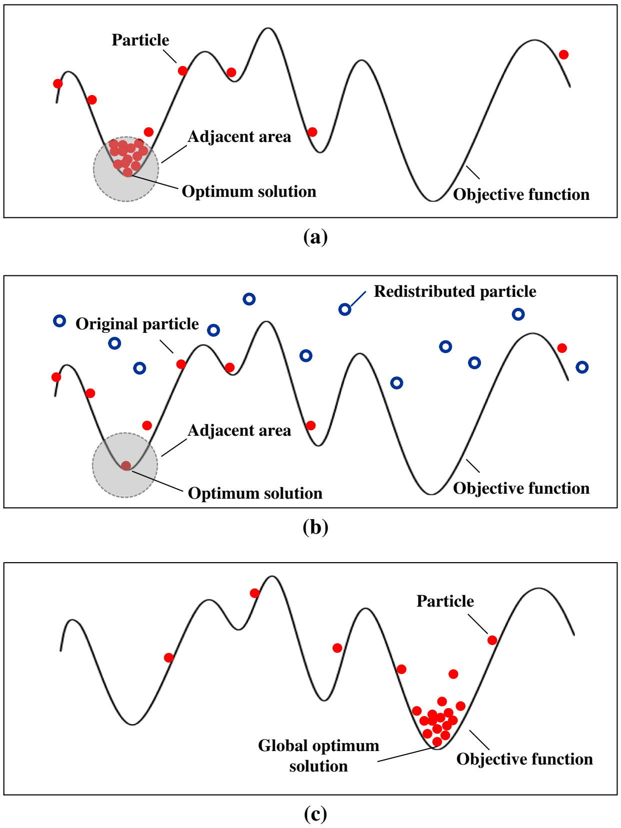

Fig. 2. Illustration of the redistributing mechanism for preventing premature convergence. In (a), when we detect premature convergence, we determine the optimum solution and its adjacent area with a radius of R. In (b), we randomly redistribute those particles in this adjacent area in the search space, except for the optimum solution, and then we continue to perform the PSO algorithm. At last, after certain times of this redistributing process, the particles converge to the true global optimum solution, as shown in (c).

B. Randomized Halton Sequences for Ensuring the Uniform Distribution and Randomness of the Initial and Redistributed Particles

Initialization of particles plays an important role in population-based optimization techniques. If the swarm population does not cover the search space efficiently, it may not be able to locate the potent solution points, thereby missing the global optimum. This difficulty may be minimized to a great extent by selecting a well-organized distribution of random numbers. Since low-discrepancy sequences have better uniformity than the random number sequences, they have been used to initialize the particles [28,29]. Comparison between the distributions of the randomly generated particles and those generated with low-discrepancy sequences is illustrated in Fig. 3.

However, the situation in this paper is a little special. We not only need to ensure the uniform distribution of the initial particles, but also need to ensure the uniform distribution and randomness of the redistributed particles when we detect premature convergence and redistribute some particles. The main goal for this is that the particles can be uniformly redistributed to different positions in the search space in different redistributing processes. Here we use Halton sequences to realize the uniform distribution of the initial and redistributed particles. A simple strategy is also presented to randomize Halton sequences and guarantee the randomness of the redistributed particles. Compared to the randomized Halton sequences proposed in Ref. [30], our method is much easier to understand and implement.

The Halton sequence is generated based on the van der Corput sequence. Now we first briefly introduced the van der Corput sequence, which is a one-dimensional low-discrepancy sequence. Let $b \geq 2$ be an integer. Any integer $n \geq 0$ can be expressed as

$$n = d_0 + d_1 b + \cdots + d_j b^j, \tag{20}$$

where $d_i \in \{0, 1, \ldots, b-1\}$ for $i = 0, 1, \ldots, j$. The radical-inverse function $\phi_b(n)$ is defined as

$$\phi_b(n) = \frac{d_0}{b} + \frac{d_1}{b^2} + \cdots + \frac{d_j}{b^{j+1}}. \tag{21}$$

The van der Corput sequence in base b is the sequence $\{\phi_b(n)\}|_{n=0}^{\infty}$.

The Halton sequence is an extension of the van der Corput sequence to a multi-dimensional space. The t-dimensional Halton sequence X_n in $(0, 1]^t$ is usually defined as

$$X_n = (\phi_{b_1}(n), \phi_{b_2}(n), \ldots, \phi_{b_t}(n)), \tag{22}$$

where $b_1, b_2, \ldots, b_t$ are integers that are greater than one and pair-wise prime. In practice, they are usually chosen to be the first t prime numbers.

Now we begin to discuss how to randomize the s-dimensional Halton sequence. The basic principle adopted here is to randomly rearrange the order of the $\phi_{b_1}(n), \phi_{b_2}(n), \ldots, \phi_{b_t}(n)$ in X_n ($b_1, b_2, \ldots, b_t$ are the first t prime numbers in order, and t is usually more than five for the case of the PD technique). This problem can be boiled down to generating a random permutation of the sequence $1, 2, \ldots, t$. Here, the Fisher–Yates shuffle algorithm is introduced to complete this work. The Fisher–Yates shuffle is an algorithm for generating a random permutation of a finite sequence (in plain terms, the algorithm shuffles the sequence). The algorithm effectively puts all the elements into a hat; it continually determines the next element by randomly drawing an element from the hat until

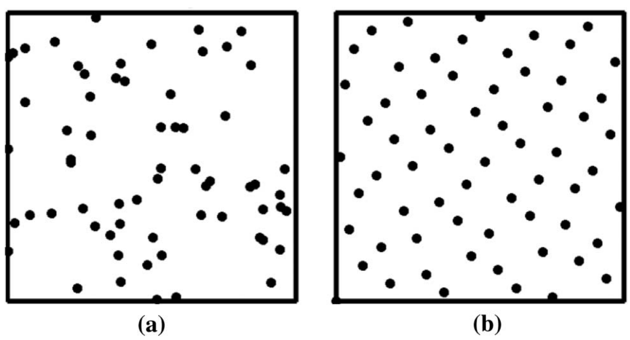

Fig. 3. Comparison between the distributions of the (a) randomly generated particles and (b) those generated with low-discrepancy sequences. We can see that the particles generated with low-discrepancy sequences have better uniformity.

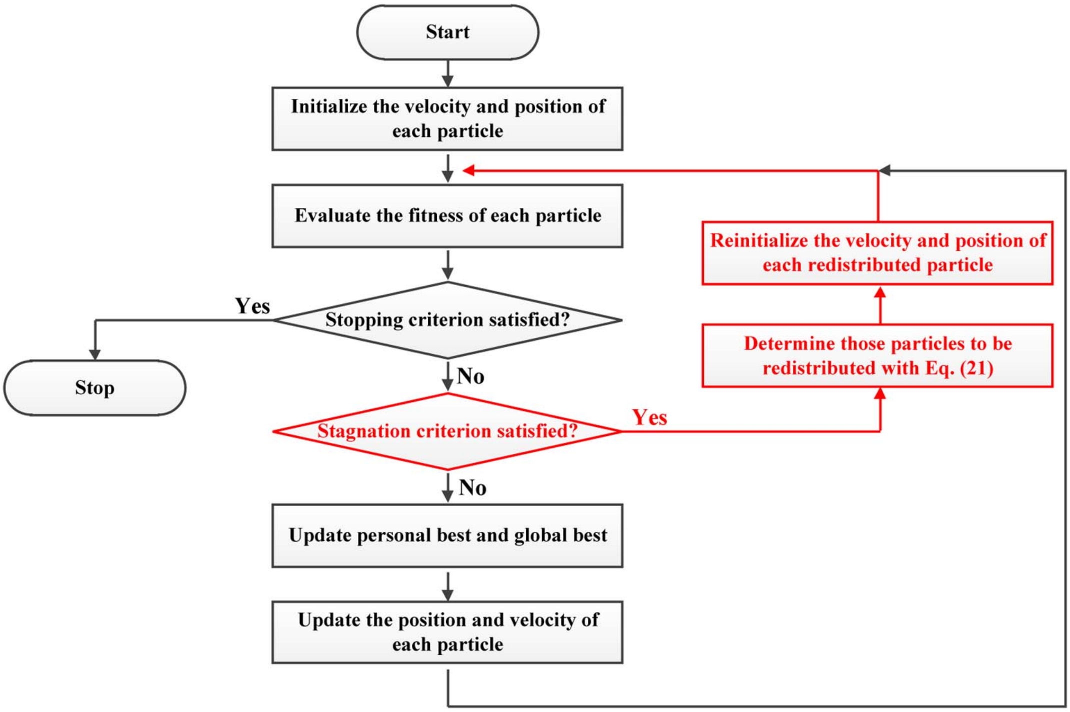

Fig. 4. Flow chart of the modified PSO algorithm for PD with an explicit strategy (in red color) for solving the stagnation problem.

no elements remain. The algorithm produces an unbiased permutation: every permutation is equally likely.

The pseudo code of the modern Fisher–Yates shuffle algorithm to shuffle an array **q** of N elements (indices $0, 1, \ldots, N-1$ is shown below:

$$\begin{aligned}&\textbf{for } i \quad \textbf{from} \quad 0 \textbf{ to } N-2 \quad \textbf{do}\\ &\quad j \leftarrow \text{random integer such that} \quad i \le j < N\\ &\quad \text{exchange } \mathbf{q}[i] \quad \text{and} \quad \mathbf{q}[j].\end{aligned} \tag{23}$$

With the Fisher–Yates shuffle algorithm presented in Eq. (23), we can obtain a random permutation of the sequence $1, 2, \ldots, t$. Using this permutation sequence, we can correspondingly set the order of $\phi_{b_1}(n), \phi_{b_2}(n), \ldots, \phi_{b_t}(n)$ in X_n, and a randomized Halton sequence is generated. Besides, we can further randomize the sequences by randomly choosing the start of each sequence.

We should emphasize that here the randomized Halton sequences are used to complement the redistributing mechanism presented above for improving the global searching ability and solving the stagnation problem of the PSO algorithm. They do not make much sense when they are discussed individually (without the redistributing mechanism).

The flow chart of the modified PSO algorithm for PD proposed in this paper is shown in Fig. 4. We should point out that when we initialize the velocity and position of the initial and redistributed particles, we ensure only the uniformity of the positions of them. The velocities of them are still randomly generated. The stagnation criterion in this figure indicates Eqs. (15) or (17).

5. SIMULATIONS AND EXPERIMENT

In this section, simulations and an experiment will be performed to demonstrate the effectiveness of the proposed strategy for solving the stagnation problem of the traditional PSO algorithm for PD.

A. Simulations

In this part, we will first perform similar simulations as in Section 3 using the modified PSO algorithm. Three different cases will be considered, which are presented below:

(1) In the first case, the range of aberration coefficients is $[-0.5\lambda, 0.5\lambda](\lambda = 500\ \text{nm})$, and the dimension is 5.

(2) In the second case, the range is increased to $[-1.0\lambda, 1.0\lambda]$, and the dimension is still 5.

(3) In the third case, the range is $[-1.0\lambda, 1.0\lambda]$, while the dimension is increased to 8.

The noise conditions considered in this part are the same as those in Section 3. Then we can use the modified PSO algorithm to recover the wavefront coefficients. Since in this algorithm there is an explicit strategy to solve the stagnation problem (thus greatly improving the global searching ability), we can further change the value of c_1 and c_2 to improve the local searching ability of the algorithm. In the modified algorithm, we set the learning factors $c_1 = c_2 = 1.05$. The value of w is constant, $w = 0.37$, $T = 0.001$, and $R = 0.05\lambda$. The population size is 40 and the maximum number of iterations is 400. The accuracy of the three cases [evaluated with Eq. (14)] considered in this part is shown in Fig. 5.

Comparing Fig. 5 with Fig. 1, we can clearly see that the accuracy of the PSO algorithm for PD is not sensitive to

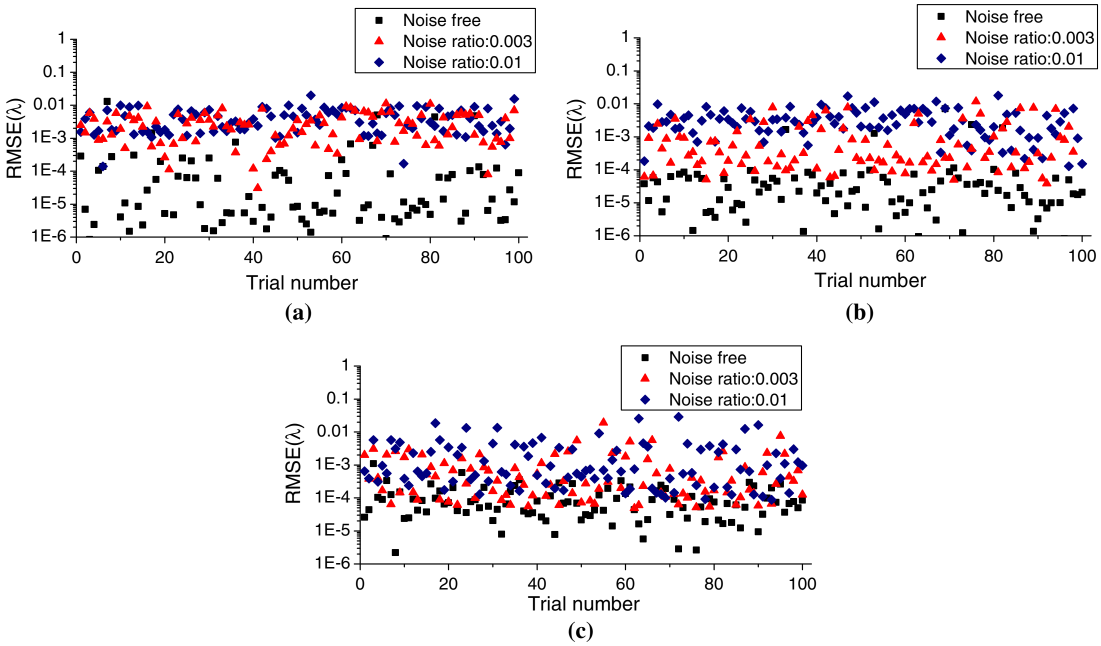

Fig. 5. Results of the modified PSO algorithm for PD under different conditions. In (a), the range of aberration coefficients to be searched is $[-0.5\lambda, 0.5\lambda](\lambda = 500$ nm), and the dimension is 5. In (b), the range is increased to $[-1.0\lambda, 1.0\lambda]$, and the dimension is still 5. In (c), the range is $[-1.0\lambda, 1.0\lambda]$, while the dimension is increased to 8. Different noise conditions are also considered in each case. We can recognize that with the strategy for solving the stagnation problem, the modified PSO algorithm for PD is no longer sensitive to noise or the range and dimension of the search space.

the range and dimension of the search space. With the increase in the range and dimension of aberration coefficients to be searched, the accuracy nearly stays unchanged. While the accuracy actually decreases with the increase in the noise ratio, it is still acceptable, and the algorithm is seldom trapped in a local minimum (in this case, a wrong result will be obtained). We can conclude that the strategy of solving the stagnation problem of the traditional PSO for PD is efficient.

To further help grasp the specific operation mechanism of the proposed strategy for solving the stagnation problem, we continue to perform some simulations to show the trend of the change of error metric in the optimization process. For one specific case (coefficient range of $[-0.5\lambda, 0.5\lambda]$, dimension of 5, and a noise ratio of 0.01), we use the traditional PSO algorithm and the modified PSO algorithm to optimize the error metric (each with four simulations), respectively. The change of the error metrics during optimization are shown in Fig. 6, and different colors in this picture are used to differentiate the results of the four simulations.

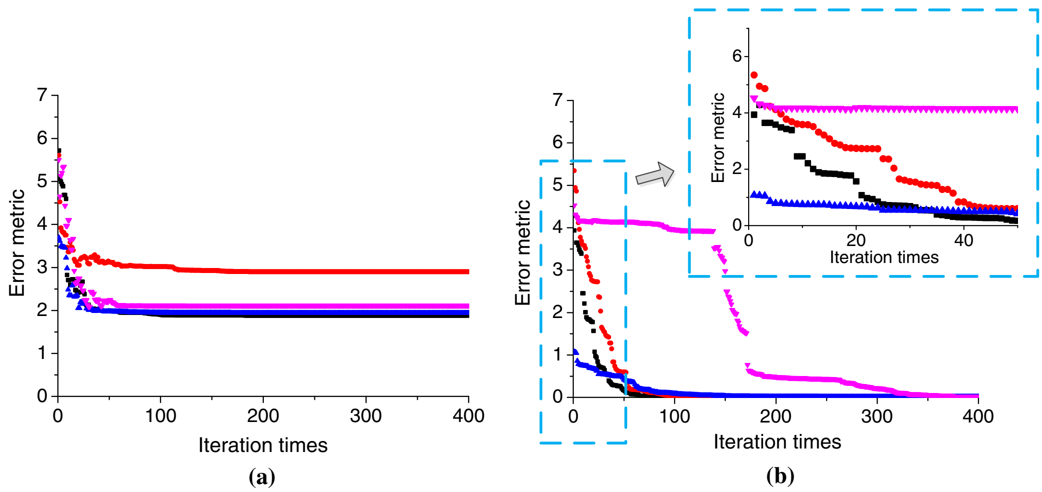

Fig. 6. Change of the error metrics (four sets of data) of the traditional PSO (a) and the modified PSO (b) for PD during the optimization process. We can clearly recognize a step-change downward trend in (b). This rapid stepwise decrease in error metric of (b) illustrates the effectiveness of the redistributing mechanism.

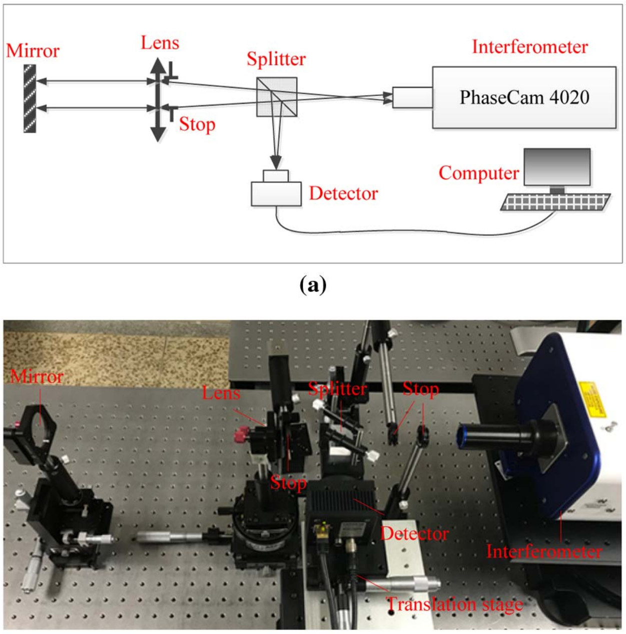

Fig. 7. (a) Sketch and (b) physical map of the optical system used in the experiment.

We can see from Fig. 6(a) that all of the four simulations using the traditional PSO algorithm are trapped in a local minimum, while all of the four simulations using the modified PSO algorithm can find the true global minimum, as shown in Fig. 6(b). Particularly, we can recognize that the error metrics of the four simulations in Fig. 6(b) change in a stepwise decrease mode. The main reason that results in this characteristic step-change decrease is the redistributing strategy. The redistributing strategy provides a mechanism for jumping out of the local minimum and thus can prevent premature convergence. When the redistributed particles find a solution better than the current optimum solution, the step-change decrease will happen, until the true global optimum solution is obtained.

Also, we can see that for most cases in Fig. 6(b), the step-change decrease happens very rapidly and the optimization process can converge to the global minimum very soon. One reason for this is the randomized Halton sequences for ensuring the uniform distribution of the redistributed particles in the search space, which can make the redistributed particles cover the search space more efficiently, thus improving the effectiveness of the redistributing mechanism.

B. Experimental Tests

To further validate the effectiveness of the proposed strategy for solving the stagnation problem of the traditional PSO

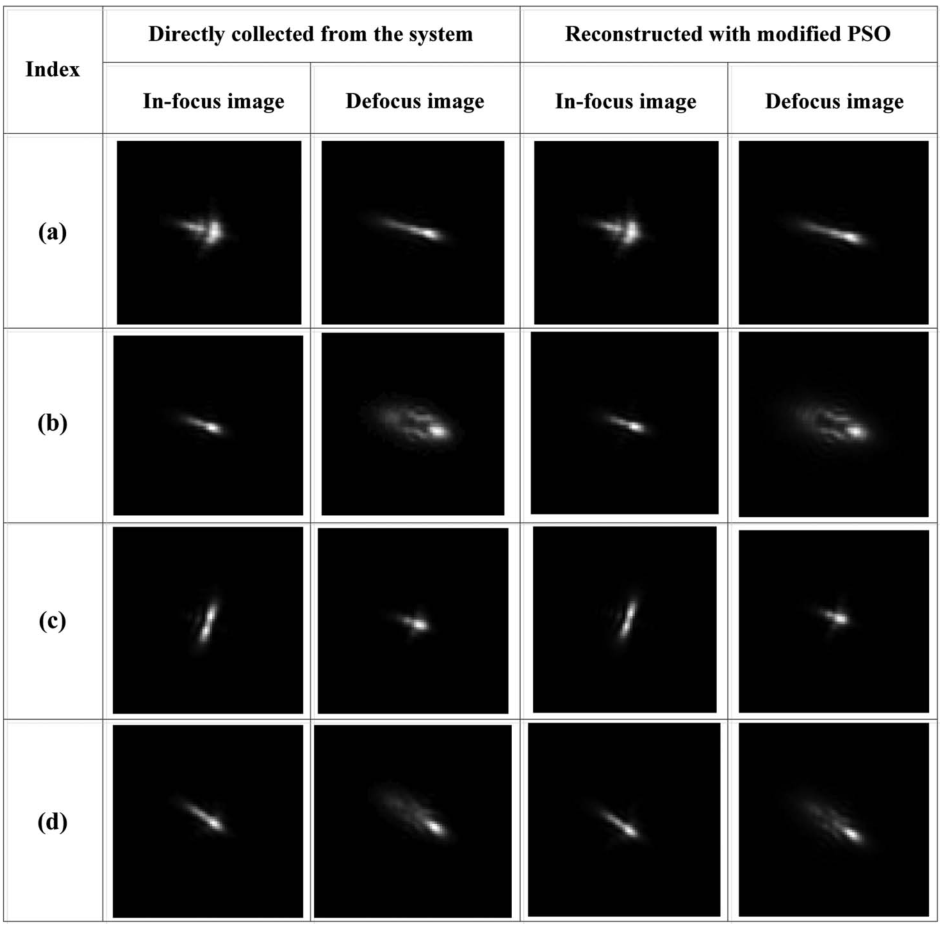

Fig. 8. Four pairs of PSF images directly collected from the optical system and re-generated with the recovered aberration coefficients.

algorithm for PD, a simple experiment will be performed. The sketch and physical map of the optical system used in the experiment are shown in Fig. 7(a) and Fig. 7(b), respectively. The interferometer (PhaseCam 4020) in Fig. 7 serves two purposes. On one hand, it can directly measure the aberrations of the optical system, which is composed of only one lens (the aberrations of this system are introduced by slightly rotating the lens). On the other hand, the focus of the interferometer (PhaseCam 4020) is used to serve as the point light source. The beam passes through the system two times and a PSF can be obtained at the detector. By using the one-dimensional precision translation stage, we can obtain a pair of PSFs with a known defocus diversity between them. Then by using PD with the modified PSO algorithm serving as the optimization method, we can recover the corresponding aberration coefficients (note that the aberration coefficients here are twice as much as the aberration coefficients of the optical system, for the beam passes through the system twice). The focal length of the lens is 180 mm, the diameter of the aperture stop is 9.2 mm, the defocus distance is 1.5 mm, the wave length is 0.6328 μm, and the pixel size of the detector is 5.5 μm.

Here, we use two methods to demonstrate the accuracy of the recovered aberration coefficients. On one hand, we can use the recovered aberration coefficients to re-generate the in-focus and defocus PSF images with Fourier optics. By comparing these two re-generated PSF images with those directly obtained from the optical system, we can validate the effectiveness of the proposed algorithm qualitatively. On the other hand, we can directly use the aberration coefficients measured by the interferometer to quantitatively demonstrate the accuracy of the proposed algorithm. Four pairs of PSF images are collected, and the corresponding aberration coefficients are measured by the interferometer and recovered by the proposed algorithm. The results are shown in Fig. 8 and Table 1, respectively.

We can see from Fig. 8 that the PSF images directly collected from the optical system bears strong similarities with those generated with the aberration coefficients recovered using the modified algorithm. This fact can qualitatively demonstrate the effectiveness of the modified algorithm. In this figure, the words "in-focus" and "defocus" mean only that the two images are collected at different focal planes.

Table 1. Aberration Coefficients Measured by Interferometer (A) and Recovered Using PD with Modified PSO Algorithm Serving as Optimization Method (B) for Four Pairs of PSF Images[a]

		C_5	C_6	C_7	C_8
(a)	A	0.621	−0.339	0.137	−0.021
	B	0.608	−0.345	0.130	−0.026
(b)	A	0.395	−0.278	0.112	−0.026
	B	0.382	−0.288	0.104	−0.033
(c)	A	0.345	−0.257	0.102	−0.020
	B	0.336	−0.264	0.095	−0.023
(d)	A	0.574	−0.264	0.127	−0.016
	B	0.561	−0.269	0.124	−0.020

[a]The aberration coefficients are measured in $\lambda = 0.6328$ μm.

The aberration coefficients directly measured from the interferometer and those recovered with the modified algorithm are shown in Table 1. Here only the 5th–8th fringe Zernike coefficients are compared, which correspond to astigmatic and coma aberrations. The defocus aberration cannot be compared. The high-order aberrations are very small, and we still do not compare them. We can see from Table 1 that the mean error of each aberration coefficient is about $0.01\lambda(\lambda = 0.6328\ \mu\text{m})$, which is enough to demonstrate the feasibility of the modified algorithm in practical conditions. The main reason for this error is the non-common path error induced by the splitter.

Also, the traditional PSO algorithm is used to recover the aberration coefficients. However, we find that it can easily be trapped in a local minimum, and a wrong result will be obtained. This result is consistent with that shown in the simulation process. In this part, we no longer present the wrong results obtained by the traditional PSO algorithm.

6. CONCLUSION

One key point of the PD technique is to develop a suitable optimization algorithm to minimize the error metric (objective function) and search the coefficient set for which the error metric is a global minimum. Gradient-based optimization algorithms can easily be trapped in a local minimum, for the searching direction of them mainly depends on the derivative information of the error metric. From this perspective, population-based optimization algorithms are better, for they rely directly upon objective function values rather than derivative information. Compared to other population-based optimization algorithms, the PSO algorithm has simple structure, high convergence efficiency, and fast searching ability. However, we should note that the traditional PSO algorithm for PD still suffers from the stagnation problem (premature convergence). This problem will become even more severe with the increase in the noise level and the scale and dimension of the aberration coefficients set to be searched.

To solve the stagnation problem of PSO for PD and improve the robustness of PD wavefront sensing, an explicit strategy is proposed in this paper. According to the inherent optimization mechanism of the PSO algorithm, we know that there exists an information exchange among different particles, and the particle swarm will gravitate towards the optimum solution (may be a local minimum), instead of randomly looking for the solution individually. Based on this basic knowledge, we propose a redistributing mechanism to detect and prevent premature convergence. Randomized Halton sequences are further introduced to complement this redistributing mechanism and improve the efficiency of it, which can ensure the uniform distribution and randomness of the initial particles and the redistributed particles in the search space. Simulation results show that this strategy can effectively improve the global searching ability of the PSO algorithm and solve its stagnation problem, especially for large-scale and high-dimension wavefront sensing and noisy conditions. The effectiveness of the proposed method is further verified by an experiment. This work can contribute to improving the capacity and robustness of PD wavefront sensing.

Funding. National Natural Science Foundation of China (NSFC) (61705223).

REFERENCES

1. R. A. Gonsalves and R. C. Hidlaw, "Wavefront sensing by phase retrieval," Proc. SPIE **207**, 32–39 (1979).
2. R. A. Gonsalves, "Phase retrieval and diversity in adaptive optics," Opt. Eng. **21**, 829–832 (1982).
3. R. G. Paxman, T. J. Schulz, and J. R. Fienup, "Joint estimation of object and aberrations by using phase diversity," J. Opt. Soc. Am. A **9**, 1072–1085 (1992).
4. R. G. Paxman, B. J. Thelen, and J. H. Seldin, "Phase-diversity correction of turbulence-induced space-variant blur," Opt. Lett. **19**, 1231–1233 (1994).
5. R. G. Paxman and J. R. Fienup, "Optical misalignment sensing and image reconstruction using phase diversity," J. Opt. Soc. Am. A **5**, 914–923 (1988).
6. D. J. Lee, M. C. Roggemann, B. M. Welsh, and E. R. Crosby, "Evaluation of least-squares phase-diversity technique for space telescope wave-front sensing," Appl. Opt. **36**, 9186–9197 (1997).
7. M. R. Bolcar and J. R. Fienup, "Sub-aperture piston phase diversity for segmented and multi-aperture systems," Appl. Opt. **48**, A5–A12 (2009).
8. I. Klapp and J. Rosen, "Phase diversity implementation in Fresnel incoherent holography," in *Imaging and Applied Optics* (Optical Society of America, 2013), paper CTh3C.3.
9. N. Védrenne, L. M. Mugnier, V. Michau, M. T. Velluet, and R. Bierent, "Laser beam complex amplitude measurement by phase diversity," Opt. Express **22**, 4575–4589 (2014).
10. J. R. Fienup, "Phase retrieval algorithms: a comparison," Appl. Opt. **21**, 2758–2769 (1982).
11. J. R. Fienup, "Phase-retrieval algorithms for a complicated optical system," Appl. Opt. **32**, 1737–1746 (1993).
12. X. Rondeau, E. Thiébaut, M. Tallon, and R. Foy, "Phase retrieval from speckle images," J. Opt. Soc. Am. A **24**, 3354–3365 (2007).
13. P. M. Johnson, M. E. Goda, and V. L. Gamiz, "Multiframe phase-diversity algorithm for active imaging," J. Opt. Soc. Am. A **24**, 1894–1900 (2007).
14. D. Simon, "Biogeography-based optimization," IEEE Trans. Evol. Comput. **12**, 702–713 (2008).
15. R. Haupt, "Comparison between genetic and gradient-based optimization algorithms for solving electromagnetics problems," IEEE Trans. Magn. **31**, 1932–1935 (1995).
16. D. Whitley, "A genetic algorithm tutorial," Stat. Comput. **4**, 65–85 (1994).
17. K. F. Tehrani, J. Xu, Y. Zhang, P. Shen, and P. Kner, "Adaptive optics stochastic optical reconstruction microscopy (AO-STORM) using a genetic algorithm," Opt. Express **23**, 13677–13692 (2015).
18. S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," Science **220**, 671–680 (1983).
19. S. Kirkpatrick, "Optimization by simulated annealing: quantitative studies," J. Stat. Phys. **34**, 975–986 (1984).
20. J. Kennedy and R. Eberhart, "Particle swarm optimization," in *IEEE International Conference on Neural Networks (ICNN)* (1995), pp. 1942–1948.
21. I. C. Trelea, "The particle swarm optimization algorithm: convergence analysis and parameter selection," Inf. Process. Lett. **85**, 317–325 (2003).
22. M. Clerc, "The swarm and the queen: towards a deterministic and adaptive particle swarm optimization," in *Evolutionary Computation (CEC)* (IEEE, 1999) pp. 1951–1957.
23. M. Taherkhani and R. Safabakhsh, "A novel stability-based adaptive inertia weight for particle swarm optimization," Appl. Soft Comput. **38**, 281–295 (2016).
24. K. F. Tehrani, Y. Zhang, P. Shen, and P. Kner, "Adaptive optics stochastic optical reconstruction microscopy (AO-STORM) by particle swarm optimization," Biomed. Opt. Express **8**, 5087–5097 (2017).
25. P. G. Zhang, C. L. Yang, Z. H. Xu, Z. L. Cao, Q. Q. Mu, and L. Xuan, "Hybrid particle swarm global optimization algorithm for phase diversity phase retrieval," Opt. Express **24**, 25704–25717 (2016).
26. G. Xu, Z. H. Wu, and M. Z. Jiang, "Premature convergence of standard particle swarm optimisation algorithm based on Markov chain analysis," Int. J. Wireless Mobile Comput. **9**, 377–382 (2015).
27. Q. Yuan and G. Yin, "Analyzing convergence and rates of convergence of particle swarm optimization algorithms using stochastic approximation methods," IEEE Trans. Automatic Cont. **60**, 1760–1773 (2015).
28. G. Weerasinghe, H. Chi, and Y. Cao, "Particle swarm optimization simulation via optimal Halton sequences," Procedia Comput. Sci. **80**, 772–781 (2016).
29. M. Pant, R. Thangaraj, and C. Grosan, "Improved particle swarm optimization with low-discrepancy sequences," in *IEEE Congress on Evolutionary Computation (CEC) (IEEE World Congress on Computational Intelligence)* (IEEE, 2008), pp. 3011–3018.
30. X. Wang and F. J. Hickernell, "Randomized Halton sequences," Math. Comput. Modell. **32**, 887–899 (2000).