An Improved Algorithm for Multi-Plane Reconstruction with Gerchberg-Saxton Phase Retrieval Algorithm

Li Fang[1,2] Bi Yong[2] Kong Xinxin[1,2] Wang Hao[1,2] Sun Minyuan[2] Li Tuo[1]

([1] University of Chinese Academy of Sciences, Beijing 100049, China
[2] Academy of Opto-Electronics, Chinese Academy of Sciences, Beijing 100094, China)

Abstract Gerchberg-Saxton (GS) phase retrieval iteration algorithm is widely used in multi-plane reconstruction despite of the low correlation coefficient C values of most planes. Method of compensation, which we call strong compensation, was proposed by the researchers but it has little function. In this paper, an improved algorithm based on the concept of weak compensation is proposed. Introduction of weight factors to adjust the balance of the multi-plane reconstruction and to realize the weak compensation is the core concept in this proposal. The simulation results show that with this method the average and variation of C values can be improved by at least 59.82% and 97.03% compared with original GS iterative algorithm, respectively.

Key words holography; weak compensation; numerical simulation; Gerchberg-Saxton phase retrieval algorithm; correlation value; image quality

OCIS codes 090.1760; 090.1970;090.2870

一种基于GS相位恢复算法的全息多平面显示的改进算法

李 芳[1,2] 毕 勇[2] 孔新新[1,2] 王 皓[1,2] 孙敏远[2] 李 拓[1]

([1] 中国科学院大学，北京 100049
[2] 中国科学院光电研究院，北京 100094)

摘要 Gerchberg-Saxton(GS)相位恢复算法被广泛用在多平面全息显示中，但是大部分平面的相关度很低。研究者们提出了一种补偿的方法，然而它并不有效。本文将这种补偿方法称作强补偿方法，并提出了一种基于弱补偿概念的方法。这个算法的核心是引进相应的权重因子改变进入算法的信息量从而实现弱补偿。数值模拟结果显示，同原始GS相位恢复三维算法相比，用这种算法各平面相关度的平均值和差值可以至少改善59.82%和97.03%。

关键词 全息；弱补偿；数值模拟；Gerchberg-Saxton相位恢复算法；相关度；像质

中图分类号 O438.1 **文献标识码** A **doi**: 10.3788/CJL201340.1009001

1 Introduction

Gerchberg-Saxton (GS) phase retrieval iteration algorithm[1], proposed in 1972, is an iterative phase retrieval algorithm used to calculate phase-only computer hologram in holographic projection and other areas[2]. It has been widely used due to its high calculation efficiency. For conventional holographic multi-plane reconstruction, there are three iteration methods adopted to generate the phase-only hologram for multi-plane reconstruction with GS algorithm. The first is based on consecutive Fresnel transform between hologram plane and image plane[3-5]; the second is based on the addition of holographic complex distribution calculated directly from the multiple planes[6-7]; the third generates holograms by applying the in-sequence Fresnel iteration between hologram plane and image plane for multiple planes[8]. Besides, Fresnel transform can be substituted by fast Fourier

收稿日期: 2013-05-02; **收到修改稿日期**: 2013-05-09

基金项目: 中国科学院光电研究院创新项目

作者简介: 李 芳(1988—)，女，硕士研究生，主要从事全息三维显示方面的研究。E-mail: lifang@aoe.ac.cn

导师简介: 毕 勇(1973—)，男，博士，研究员，主要从事激光器的激光显示方面的研究。E-mail: biyong@aoe.ac.cn

transform (FFT) via introducing an additional lens factor for the multiple planes in these three methods[6,9-10]. To compute Fresnel diffraction integration, there are also three methods, single FFT (SFFT), double FFT (DFFT) and angular spectrum[11]. SFFT is faster than others and can be used to generate larger images, but it is only suitable for large diffraction distance.

Multi-plane hologram calculation has drawn extensive attention of researchers. However, correlations of reconstructed images obtained by general methods mentioned above are generally low. For example, in the three. plane reconstruction case, only one correlation coefficient C can reach 0.9 or 0.8, the others are very low. In our simulation experiments applying the other two methods, the results are similar. C variation among these images is also large. Therefore, these methods are limited in practical applications.

In order to solvethose problems, many new algorithms based on the GS phase retrieval iteration algorithm have been proposed. As an example, Ying *et al.*[12] proposed an algorithm combining GS iteration and compensation iteration, which has improved correlation but with the trade-off of calculation time and increased complexity. Shi *et al.*[8] devised a method based on the concept of compensation to address the iterative sequence of multiple planes. However, it has a serious problem and our experiment indicates that when the pictures are changed, the proper sequence of the iteration for multiple planes also varies and correlations largely decrease in most situations. Because this method is influenced by the images and the proper iterative sequence, it is not stable for multi-plane reconstruction. Since the compensation is very strong in this method, we term it as strong compensation model hereafter in the discussion.

In this paper, the strong compensation model is analyzed and a new method entitled the weak compensation is proposed. Through increased correlation and decreased variation, this method can substantially improve the quality of the reconstructed image without sacrificing the computation time and the algorithm complexity. The theoretical principle of this new method is analyzed and the simulation results are presented. The main advantage of this method is simplicity. It increases the average correlation value and decreases the variation simultaneously without the need of more iterations. Besides, this algorithm is stable when there are more pictures involved and pictures changed.

2 Theory

2.1 GS phase retrieval iterative algorithm

GS phase retrieval iteration algorithm has been chosen to calculate the phase-only hologram for multi-plane reconstruction[8].

At first the phase distribution of the hologram is assumed to random numbers between 0 and 2π. Then the consecutive iterations for multiple planes to calculate the phase distribution of the hologram start. When the number of the reconstructed multiple planes is N, every one iteration consists of N loops and every loop is one GS ping-pong iteration for one plane. A single loop is expressed as

$$\begin{cases} F_{z_i}^{\lambda}[\exp(\mathrm{j}\varphi_{i+}^{k})] = g_i^k \exp(\mathrm{j}\theta^k) \\ F_{z_i}^{(-1)\lambda}[g_{0i}\exp(\mathrm{j}\theta^k)] = f^k \exp(\mathrm{j}\varphi_{i-}^{k}) \end{cases}, \tag{1}$$

where $F_{z_i}^{\lambda}$ and $F_{z_i}^{(-1)\lambda}$ denote the Fresnel transform and Fresnel inverse transform with the wavelength λ and the distance z_i; superscript k denotes the kth iteration; subscript i denotes the ith loop for ith plane; φ_{i+}^{k} and φ_{i-}^{k} denote the phase distribution of hologram when the ith loop begins and ends, respectively; θ^k is the phase distribution in the image plane. The amplitude restraint is that g_i^k is replaced by g_{0i} which is the ideal amplitude of the plane i for the ith loop. For the next loop, the phase distribution of hologram is replaced by

$$\varphi_{(i+1)+}^{k} = \varphi_{i-}^{k}. \tag{2}$$

When N loops from the first plane to the last plane are finished the kth iteration is finished. The schematic of the kth multi-plane iteration is shown in Fig. 1 (taking three-plane reconstruction for example)[11]. g1, g2 and g3 denote the three planes; 1～6 denote the Fresnel transform and Fresnel inverse transform between hologram plane and three image planes; the digits denote the sequence.

The correlation coefficient C between the iterative image and the original image is chosen to evaluate the result of the algorithm and it is the measure of the resemblance of the output image with the input image.

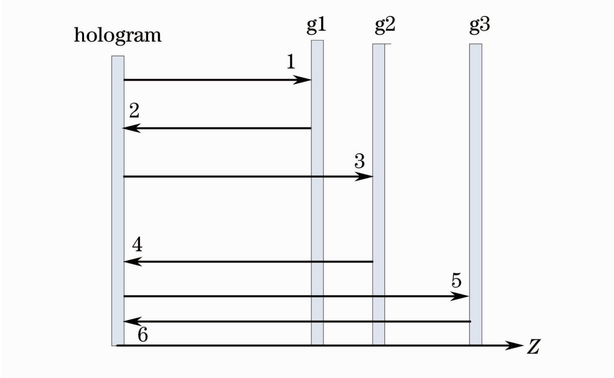

Fig. 1 Schematic of the kth multi-plane iteration.

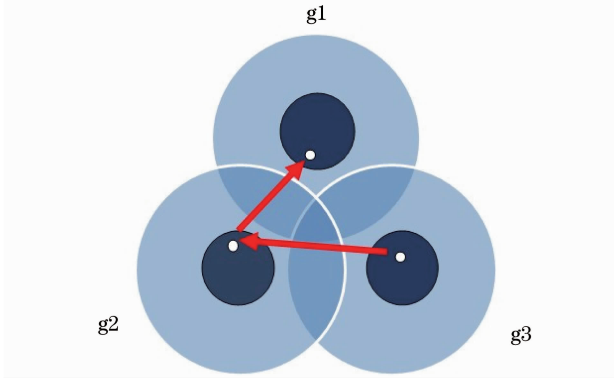

Fig. 2 Principle of strong compensation.

It is defined as

$$C(t,t_0) = \mathrm{cov}(t,t_0)(\sigma_t \cdot \sigma_{t_0})^{-1}. \tag{3}$$

where $\mathrm{cov}(t,t_0)$ is the cross-covariance between t and t_0, σ is the standard deviation, t_0 represents the original image and t represents the iterative image. The maximum value of unity means that t is perfectly correlated with t_0. Hence, $C(g_i^k, g_{0i})$ can be used to evaluate the quality of the reconstructed images.

2.2 Strong compensation

For each iteration that follows the order of g1→g2→g3 (taking N as three for example), the algorithm ends with the iteration for the plane g3 and the result shows that C is the largest for g3, which means the optimum for g3. Correspondingly, C values of g1 and g2 are not high. To increase C of g1 and g2, researchers have proposed concept of compensation and another reverse iteration of g2→g1 is performed[7]. The compensation iteration is expressed as

$$\begin{cases} F_{z_2}^{\lambda}[\exp(\mathrm{j}\varphi_{2+})] = g_2\exp(\mathrm{j}\theta) \\ F_{z_2}^{(-1)\lambda}[g_{02}\exp(\mathrm{j}\theta)] = f\exp(\mathrm{j}\varphi_{2-}) \\ F_{z_1}^{\lambda}[\exp(\mathrm{j}\varphi_{1+})] = g_1\exp(\mathrm{j}\theta) \\ F_{z_1}^{(-1)\lambda}[g_{01}\exp(\mathrm{j}\theta)] = f_1\exp(\mathrm{j}\varphi_{1-}) \end{cases}, \tag{4}$$

and $\varphi_{2-} = \varphi_{1+}$.

Since the compensation is very strong in this method, we term it as strong compensation model in the discussion. Then we will study the strong compensation method and our proposed weak compensation method from anew viewpoint of solution space. Actually, the end point of the algorithm is in the solution space and thus we can analyze the whole iterative process in the solution space. Figure 2 shows the principle of the strong compensation from the viewpoint of solution.

Three big circles in Fig. 2 represent the phase distribution solution set of GS iteration for planes g1, g2 and g3, respectively. Their intersection area represents the public solution of g1, g2 and g3, which is wanted. The phase distribution solution of hologram hops among these three solution sets during the iteration. Three small circles with dark color represent the optimum phase distribution solution set in which C values are larger and close to 1. When the consecutive iterations for g1→g2→g3 are finished, the phase distribution solution of hologram is located in the area of optimum solution set of g3 or near this area. When the compensation of iteration of g2 is performed, the solution will move to the area of optimum solution set of g2. If there is no additional treatment employed, the compensation of iteration of g2 is too strong and the solution moves far away from the solution set of g3. As a result, the C value of g2 will rise but that of g3 will drop a lot. Similarly, when the compensation of iteration of g1 is performed, the location moves to the area of optimum solution set of g1. With no additional measures the C value is large and close to 1 for the plane in which the iteration ends but for the other two planes their C values are very small. The concept of compensation is useful but ineffective to achieve optimum C value for all the three planes.

It is theoretically possible that with the strong compensation the proper sequence of iteration and compensation can be found[7], but when the pictures change, this sequence may lead to worse results compared with the situation with no compensation. Our experimental results show that the average value of C of three images actually decreases by 28.3% after the compensation. Sequence of iteration and compensation which has characteristics of good convergence has to be adjusted for each different set of target pictures.

2.3 Weak compensation

Based on the work on strong compensation method, we propose a method named weak compensation which is capable of increasing the reconstructed image quality and is significant to the field of multi-plane reconstruction.

The reverse iteration of g2→g1 is still performed but weight factors have been introduced to realize the weak compensation, with the process of compensation expressed as

$$\begin{cases} F_{z_2}^{\lambda}[\exp(\mathrm{j}\varphi_{2+})] = g_2\exp(\mathrm{j}\theta) \\ F_{z_2}^{(-1)\lambda}\{[A\times g_{02} + (1-A)\times g_2]\exp(\mathrm{j}\theta)\} = f\exp(\mathrm{j}\varphi_{2-}) \\ F_{z_1}^{\lambda}[\exp(\mathrm{j}\varphi_{1+})] = g_1\exp(\mathrm{j}\theta) \\ F_{z_1}^{(-1)\lambda}\{[B\times g_{01} + (1-B)\times g_1]\exp(\mathrm{j}\theta)\} = f_1\exp(\mathrm{j}\varphi_{1-}) \end{cases}, \tag{5}$$

and $\varphi_{2-} = \varphi_{1+}$.

In the Eq. (5), A and B are weight factors introduced to adjust the information participating in the iteration of planes g2 and g1. With the use of small values for A and B, value of $1-A$ and $1-B$ are close to 1, and hence the weak compensation is realized. Figure 3 shows the principle of weak compensation from the viewpoint of the solution space.

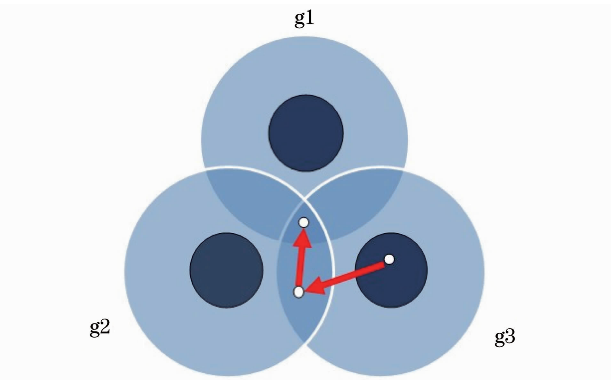

Fig. 3 Principle of weak compensation.

When the consecutive iterations for g1→g2→g3 are finished, the phase distribution solution ofhologram is still located in the area of optimum solution set of g3 or near this area. When the weak compensation of iteration for g2 is performed, with small A, the information of image g2 participating in the iteration is restricted and the solution will still be located in the phase distribution solution set of g3 but will also enter into the region of phase distribution solution set of g2, a little far away from the optimum solution circle. It means that the solution will be located in the intersection area of solution set of g2 and g3. Then, the weak compensation of iteration of g1 is performed and with the small value of B the solution will enter into the intersection region of the solution set of g1, g2, g3 which is optimum area wanted. In this way, the optimization of three planes is realized simultaneously and the values of C for three planes, as well as the average value, are high. On the other hand, with the adjustment of the values of A and B, the equalization of C for three planes can be achieved, which means that the variation is very low. The improvement in these two values helps to increase the quality of reconstruction in practical applications and thus this method makes a significant improvement.

With the proposed weak compensation method in this work, the algorithm is not sensitive to the change of pictures. For any pictures we can find the solution which is located in the intersection region of the solution set of g1, g2, g3 via constrain of the weak compensation through the adjustment of parameters A and B. Nevertheless, the adjustment of parameters A and B (the approach of adjustment see Section 4) is much more operable than searching for the proper sequence of iteration in strong compensation method, since sequence of iteration and compensation is strenuous to find and it needs vast trials. Most importantly, you might not find a proper sequence of good solution convergence.

3 Simulation results

Computer simulations of three-plane reconstruction are preformed. Figure 4 shows images of g1, g2 and g3. The images of g1, g2, g3 are all gray-level image with 256 pixel × 256 pixel size and 8 μm pixel pitch. The angular spectrum method of Fresnel transform is used in our simulation and so the computed hologram is also an image with 256 pixel × 256 pixel size and 8 μm pixel pitch. The operation wavelength is 532 nm. Three images are located at (z_1, z_2, z_3) of (1.5, 1.8, 2) (unit: m) away from the phase-only hologram. Matlab software is used for the simulation process. According to Fig.2, every kth iteration leads to similar result and it only needs one iteration of compensation carried out at the end of the algorithm to realize the improvement of results. So the result generally is not influenced by the iteration number adopted. In addition, this method has another advantage that it can achieve good result with less iterations. In our simulation experiments 50 iterations are adopted.

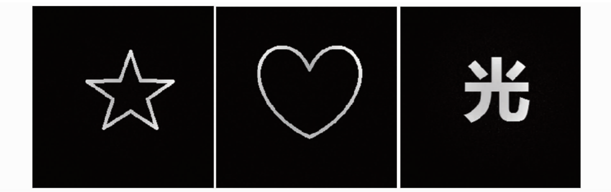

Fig.4 Images of g1, g2 and g3.

At first, the simulation with no compensation is performed and by 50 iterations the C values of three images are (0.1551, 0.3530, 0.9486) and the average C is 0.4856. The variation between the minimum and the maximum C is 0.7935. Then the simulation with strong compensation is performed and the C values become (0.9622, 0.2430, 0.0818), the average value of which becomes 0.429, decreased by 11.66%. The variation is 0.8804, decreased by 10.95%. When the weak compensation method proposed in this paper is performed (the values of A and B are 0.03 and 0.022, respectively), the C values turn to (0.7636, 0.7872, 0.7773). The average value is 0.7761, which has been improved by 59.82% compared with no compensation and 80.91% compared with strong compensation. The variation drops to 0.0236 which has been improved by 97.03% and 97.32% when compared with the result of the previous two. Figure 5 shows the reconstructed images of g1, g2, g3 and the calculated hologram with the weak compensation algorithm. Figures 6 and 7 show the reconstructed images with no compensation and with the strong compensation.

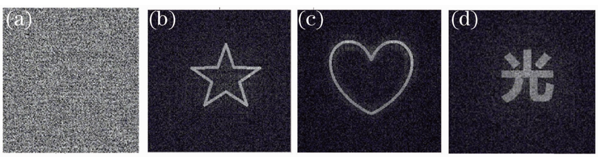

Fig.5 Reconstructed images of g1, g2 and g3 and computed hologram. (a) hologram; (b) g1; (c) g2; (d) g3.

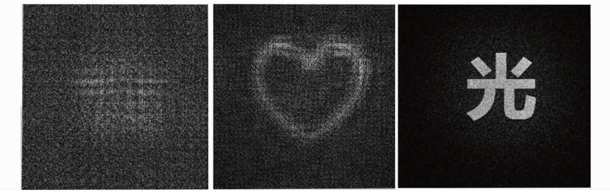

Fig.6 Images of g1, g2 and g3 with no compensation.

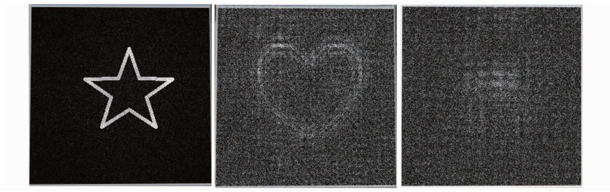

Fig.7 Images of g1, g2 and g3 with strong compensation.

Computer simulations for another group are performed. In this group, the images are expanded to 512 pixel × 512 pixel in size through the addition of black edging. The C values of images will be increased. After the simulation with weak compensation method (the values of A and B are 0.055 and 0.04, respectively), the C values of three planes become (0.8587, 0.8574, 0.8654) with average value of 0.8605 and the variation is 0.008. The average value has been improved by 83.87% compared with no compensation and 100.6% compared with the strong compensation method. The variation has been improved by 99.09% and 99.12%, respectively. The iteration number is still 50.

The third group of simulations for the 1024 pixel × 1024 pixel images is implemented. The images are the same as that of the first two group. The values of A and B are 0.13 and 0.075, respectively.

With the size of hologram increasing, the average value of C becomes higher and the weak compensation algorithm we proposed is more efficient compared with the situation with no compensation. Table 1 shows the simulation results with different hologram sizes. Table 2 shows the improvement of the C average value and variation with the weak compensation algorithm compared with no compensation and strong compensation.

The images of g1, g2, g3 are similar to the binary black-white images. To further justify the effectiveness of the algorithm for the gray-scale images, we finished the experiment with g1, g2, g3 replaced by three other gray-scale images (shown in Fig. 8) and the C values of three images are (0.7831, 0.7902, 0.8061) (A and B are 0.09, 0.0058) while the results with original GS algorithm are (0.5831, 0.6163, 0.9138) and the results with strong compensation are (0.8898, 0.6476, 0.5224). So from the results we can conclude that the strong compensation method can lead to worse results and our algorithm is still effective to gray-scale pictures.

The step-by-step description of the algorithm realized in Matlab software is given in Fig. 9.

Fig. 8 Three grayscale images.

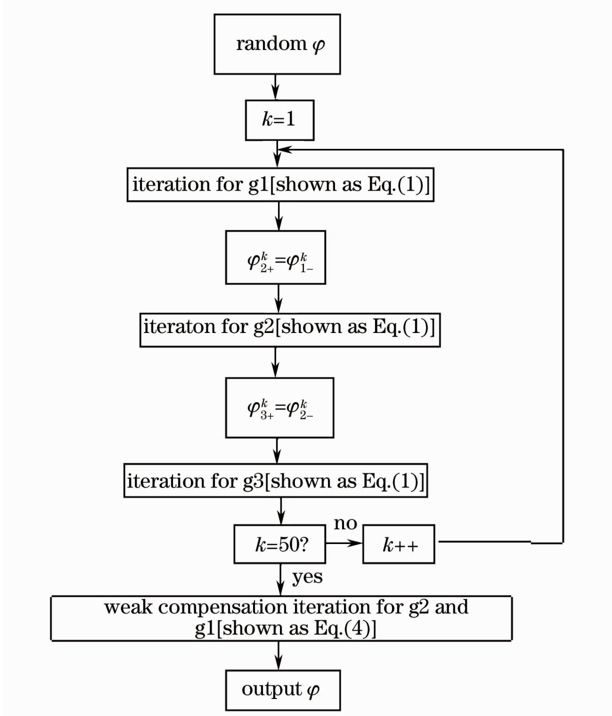

Fig. 9 Step-by-step description of the algorithm realized in Matlab.

Table 1 Simulation results with the differenthologram sizes

Image size /pixel	Compensation	C values				
		g1	g2	g3	Average	Difference
256×256	without compensation	0.1551	0.3530	0.9486	0.4856	0.7935
	strong compensation	0.9622	0.2430	0.0818	0.429	0.8804
	weak compensation	0.7636	0.7872	0.7773	0.7761	0.0236
512×512	without compensation	0.0929	0.3433	0.9679	0.4680	0.875
	strong compensation	0.9521	0.2971	0.0391	0.429	0.913
	weak compensation	0.8587	0.8547	0.8654	0.8605	0.008
1024×1024	without compensation	0.0556	0.3550	0.9564	0.4557	0.9008
	strong compensation	0.9594	0.3355	0.0501	0.4483	0.9093
	weak compensation	0.8865	0.8660	0.8703	0.8743	0.0162

Table 2 Improvement of C average value andvariation with the weak compensation algorithm compared with no compensation and strong compensation

Image size /pixel	Comparison	Improvement of average /%	Improvement of variation /%
256×256	with no compensation	59.82	97.03
	with strong compensation	80.91	97.32
512×512	with no compensation	83.87	99.09
	with strong compensation	100.6	99.12
1024×1024	with no compensation	91.86	98.20
	with strong compensation	95.03	98.22

4 Discussion

Two weight factors A and B are introduced in our algorithm to implement the weak compensation. When the images are changed, adjustment of values of A and B is needed. With the proper selection of A and B, the equalization of the C values of three reconstructed images is achieved. In this section, we will discuss the selection of the weight factors. A and B represent the information amount of images g2 and g1 to participate in the iteration. For example, if C of g3 is higher than that of g2 and g1, increase of A and B is expected and *vice versa*. If C of g2 is higher than that of g1, increase of B and decrease of A are expected and *vice versa*. In this way, we will obtain the proper values of the weight factors. To realize the weak compensation, the values of A and B are mostly very small (0.03 and 0.022 in the 256 pixel × 256 pixel simulation for example).

After the weak compensation is implemented, when the images are changed, the weak compensation is still effective except adjustment of values of A and B to realize the equalization of C values of three planes, with the method discussed. Therefore, the weak compensation method proposed in this paper is stable for different systems as well as for the increasing number of reconstructed images.

A and B would have to be changed even more for a substantially different set of target images but the whole process is rule-based and mostly needs 3～4 trails. However, according to our experiments, for the strong compensation method, there will be 36 likely iteration sequences and the more reconstructed planes, the more likely iteration sequences. Furthermore, there might not be optimal sequence existing to realize the improvement of the algorithm. So the strong compensation method is not applicable. For the weak compensation method, the 50-iteration is likely to run multiple times but the number repeated is small (3～4) and when we find the proper values of A and B, the improvement is realized and the program can be run without more iterations than original GS phase retrieval algorithm. So it does not sacrifice the iteration numbers.

When the image planes are re-ordered, for example, the images of g1 and g2 are exchanged, the values of A and B are totally changed, so one must choose the values again according the approach mentioned above.

5 Conclusion

A novel method based on the concept of compensation is proposed for multi-plane reconstruction, which is termed as the weak compensation method. The theoretical principle of this method has been examined and computer simulations performed. The results show that with the proposed method, the increase of C average value, as well as the equalization of the C values of multiple planes, can be achieved. The average value and the variation of C have been improved by at least 59.82% and 97.03%. The advantage of this new method is that it can improve the two values without sacrificing the calculation time and algorithm complexity. Besides, it is stable for different pictures and increasing number of images and can be widely used in practical applications, such as holographic display, multiple-image hiding, diffractive optical element designing and so on.

References

1 R W Gerchberg, W O Saxton. A practical algorithm for the determination of phase from image and diffraction plane pictures [J]. Optik, 1972, 35(2): 237－246.

2 Tao Tao, Jing Li, Qian Long, *et al.*. 3D trapping and manipulation of micro-particles using holographic optical tweezers with optimized computer-generated holograms [J]. Chin Opt Lett, 2011, 9(12): 120010.

3 M Makowski, M Sypek, A Kolodziejczyk, *et al.*. Three-plane phase-only computer hologram generated with iterative Fresnel algorithm [J]. Opt Eng, 2005, 44(12): 125805.

4 M Makowski, M Sypek, A Kolodziejczyk, *et al.*. Iterative design of multi-plane holograms experiments and applications [J]. Opt Eng, 2007, 46(4): 045802.

5 J Xia, H Yin. Three-dimensional light modulation using phase-only spatial light modulator [J]. Opt Eng, 2009, 48(2): 020502.

6 T Haist, M Schonleber, H J Tiziani. Computer-generated holograms from 3D-objects written on twisted-nematic liquid crystal displays [J]. Opt Commun, 1997, 140(4-6): 299－308.

7 G Sinclair, J Leach, P Jordan, *et al.*. Interactive application in holographic optical tweezers of a multi-plane Gerchberg-Saxton algorithm for three dimensional light shaping [J]. Opt Express, 2004, 12(8): 1665－1670.

8 Y Shi, G Situ, J Zhang. Multiple-image hiding in the Fresnel domain [J]. Opt Lett, 2007, 32(13): 1914－1916.

9 H Zhang, J Xie, J Liu, *et al.*. Elimination of a zero-order beam induced by a pixilated spatial light modulator for holographic projection [J]. Appl Opt, 2009, 48(30): 5834－5841.

10 I Ducin, K Kakarenko, M Makowski, *et al.*. Holographic color projection with additional phase factor to suppress zero diffractive order [C]. SPIE, 2011, 7746: 774601.

11 J W Goodman. Introduction to Fourier Optics (3rd ed) [M]. New York: McGraw-Hill, 2005. 32－45.

12 C Ying, H Pang, C Fan, *et al.*. New method for the design of a phase-only computer hologram for multi-plane reconstruction [J]. Opt Eng, 2011, 50(5): 055802.

栏目编辑：宋梅梅