sensors

MDPI

Article

Gradient-Descent-like Ghost Imaging

Wen-Kai Yu *, Chen-Xi Zhu, Ya-Xin Li, Shuo-Fei Wang and Chong Cao

Key Laboratory of Advanced Optoelectronic Quantum Architecture and Measurement of Ministry of Education, School of Physics, Beijing Institute of Technology, Beijing 100081, China; 3120191480@bit.edu.cn (C.-X.Z.); 3120181429@bit.edu.cn (Y.-X.L.); 3120195769@bit.edu.cn (S.-F.W.); 3120191424@bit.edu.cn (C.C.)

* Correspondence: yuwenkai@bit.edu.cn

Abstract: Ghost imaging is an indirect optical imaging technique, which retrieves object information by calculating the intensity correlation between reference and bucket signals. However, in existing correlation functions, a high number of measurements is required to acquire a satisfied performance, and the increase in measurement number only leads to limited improvement in image quality. Here, inspired by the gradient descent idea that is widely used in artificial intelligence, we propose a gradient-descent-like ghost imaging method to recursively move towards the optimal solution of the preset objective function, which can efficiently reconstruct high-quality images. The feasibility of this technique has been demonstrated in both numerical simulation and optical experiments, where the image quality is greatly improved within finite steps. Since the correlation function in the iterative formula is replaceable, this technique offers more possibilities for image reconstruction of ghost imaging.

Keywords: ghost imaging; gradient-descent; image reconstruction; iteration; image quality; denoising

check for updates

Citation: Yu, W.-K.; Zhu, C.-X.; Li, Y.-X.; Wang, S.-F.; Cao, C. Gradient-Descent-like Ghost Imaging. *Sensors* **2021**, *21*, 7559. https://doi.org/10.3390/s21227559

Academic Editor: Min Yong Jeon

Received: 20 September 2021
Accepted: 9 November 2021
Published: 13 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Ghost imaging (GI) [1] is an imaging technique that produces the object image via the intensity correlation of two beams: one beam interacts with the target but is collected by a bucket detector without spatial resolution (only recording the total intensity); the other beam records the spatial distribution of the light field of the source but never interacts with the target. The detector in either beam cannot "see" the object by itself, but the object image can be recovered by calculating the intensity correlation between these two arms' signals [2–4]. Later, it was found that the reference arm containing the spatially resolved array detector can be removed by using a programmable spatial light modulator (SLM) encoded with preset modulated patterns [5,6].

Over the last two decades, GI has attracted a lot of attention [7–11] and has been used in many fields, such as microscopic imaging [12], optical encryption [13], and cryptographic key distribution [14]. However, the signal-to-noise ratios (SNRs) of conventional GI methods are extremely low, even under a large number of measurements, and increasing the number of measurements only provides a limited improvement in image quality. To improve the image quality and imaging efficiency of the GI, many reconstruction algorithms have been proposed, such as background-removal GI (BGI) [15], high-order GI (HGI) [16,17], differential GI (DGI) [18], pseudo-inverse GI (PGI) [19], compressive GI [20–22], adaptive compressive GI [23], iterative GI [24–26], Gerchberg–Saxton-like GI [27], iterative eigenmode GI [28], joint iteration GI [29], singular value decomposition GI [30], etc. However, these methods have their limitations. The BGI finds it hard to deal with quasitransparent complex objects; the HGI relies on the increase in exponential power to improve visibility; the DGI can provide a considerable but relatively limited SNR improvement, which is mainly dependent on the object [27,31]; the PGI generally takes a long time to reconstruct large-scale object images and is sensitive to noise. The compressive GI algorithms require the sparse priors of objects and take huge matrix calculations; the

adaptive compressive GI aims to reduce the sampling rate from the perspective of the object's hierarchical structure, but has high requirements for modulated pattern design. Differing from the compressive GI method and its variants, the iterative GI, Gerchberg–Saxton-like GI, iterative eigenmode GI, joint iteration GI methods do not need to rely on the sparse priors of images, and can finish the image reconstruction task only using iterations based on statistical averages, without large-scale matrix multiplication, which undoubtedly greatly reduces computational consumption. However, these iterative variants of GI either have strict constraints or make some loose approximations of the expression of noise. In addition to these, researchers have made other attempts, such as singular value decomposition GI, but still need time-consuming calculations. Therefore, it is very necessary to construct a high-efficiency, high-quality image reconstruction method.

In this paper, inspired by the idea of classic gradient descent algorithms [32–35], we propose a new image reconstruction method, called gradient-descent-like ghost imaging (GGI). It is worth noting that the aim of the gradient descent method is to find the solution, which minimizes the objective function by updating the parameters in the opposite direction of the gradient of the objective function [32]. In addition, in view of the previous analysis, the correlation function (statistical average function) has a natural advantage as an iterative carrier, and we also know that, in GI, any form of intensity correlation functions can be regarded as a transformation function performed on the original image. Based on the above ideas, we gradually search for the optimal solution of the rewritten objective function to acquire high-quality image reconstruction. Both simulation and optical experiments will be performed to verify the performance of this proposed method against the noise without any sparse prior knowledge of the object images. The performance of this method in the recovery of complex object images will also be investigated.

2. Principle of Gradient-Descent-like Ghost Imaging

In this section, we will first recall the theory of gradient descent, then briefly review the common intensity correlation functions in GI and derive their matrix expression forms, and finally introduce the principle of our GGI.

2.1. Gradient Descent Theory

To our knowledge, gradient descent (also known as steepest descent) is a popular strategy, which is widely used in machine learning and deep learning to solve both convex and non-convex problems. Its idea is to gradually minimize an objective function, parameterized by a model's parameters through iterations along the opposite direction of the gradient of objective function [32].

Let J be a function with respect to an independent vector θ, denoted as $J(\theta)$, and the gradient at θ_0 can be written as $\nabla J(\theta_0)$, where ∇ denotes the gradient operator. The opposite direction of the gradient direction is also called the gradient flow of the variable θ. Assume $\theta - \theta_0 = \eta\vartheta$, where η is a positive scalar, ϑ is the unit vector of $\theta - \theta_0$, defined as $\vartheta = -\frac{\nabla J(\theta_0)}{\|\nabla J(\theta_0)\|}$. Then we can obtain $\theta = \theta_0 + \eta\left(-\frac{\nabla J(\theta_0)}{\|\nabla J(\theta_0)\|}\right) = \theta_0 - \alpha \cdot \nabla J(\theta_0)$, which is the core iterative expression of gradient descent. To obtain the minimum value of $J(\theta)$, the process of gradient descent can be performed via the following steps:

(1) Compute the current gradient (partial derivative) $\nabla J(\theta)$ with respect to θ;
(2) Multiply the current gradient by a step size (i.e., learning rate) α, i.e., $\alpha \cdot \nabla J(\theta)$, and update the variable θ via $\theta^{\text{updated}} = \theta - \alpha \cdot \nabla J(\theta)$;
(3) Repeat Steps (1–2) until the difference between the values obtained from two adjacent iterations is small enough (less than the preset termination threshold ε); at this moment, the objective function reaches its minimum;
(4) Output the current independent variable θ, which is exactly the value that minimizes the function $J(\theta)$.

For a linear measurement model, there is a hypothesis function

$$h_\theta(a_1, a_2, \cdots, a_n) = a_1\theta_1 + a_2\theta_2 + \cdots + a_n\theta_n = \Sigma_{i=1}^{n} a_i\theta_i, \tag{1}$$

where θ_i $(i = 1, 2, 3, \ldots, n)$ is the model parameter to be evaluated, a_i is the coefficient or weight. To evaluate the fitting of the algorithm, a loss function can generally be used. Minimizing the loss function will help us acquire the best fitting, and the corresponding model parameters are the optimal solution. In linear regression, the loss function is usually the square of the difference between the hypothesis function and the sample output. For the convenience of derivation, we use a loss function defined as half of the mean square error (MSE):

$$J(\theta_1, \theta_2, \cdots, \theta_n) = \frac{1}{2}\text{MSE} = \frac{1}{2m}\Sigma_{j=1}^{m}(h_\theta(a_1^{(j)}, a_2^{(j)}, \cdots, a_n^{(j)}) - y_j)^2, \tag{2}$$

where y_j denotes the jth actual measured value. Then, we calculate the partial derivative of J with respect to θ_i:

$$\frac{\partial}{\partial\theta_i}J(\theta) = \frac{1}{m}\Sigma_{j=1}^{m}(\Sigma_{i=1}^{n} a_i\theta_i^{(j)} - y_j))a_i^{(j)}. \tag{3}$$

Thus, the iterative expression can be rewritten as

$$\theta_i^{\text{updated}} = \theta_i - \alpha \cdot \frac{1}{m}\Sigma_{j=1}^{m}(\Sigma_{i=1}^{n} a_i\theta_i^{(j)} - y_j)a_i^{(j)}, \tag{4}$$

which is also called batch gradient descent [32] because it uses the gradient data of all samples when calculating the gradient.

Since each a_i $(i = 1, 2, 3, \ldots, n)$ can also be a column vector, the hypothesis function can be written as $h_\theta(A) = A\theta$, where $h_\theta(A)$ is a column vector of $m \times 1$ consisting of expected measurements, θ is a column vector of $n \times 1$ to be reconstructed, and A is a matrix of $m \times n$, which can be written as $A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$. Next, we will introduce the matrix representation of gradient descent. The foregoing loss function can be rewritten in matrix form as

$$J(\theta) = \frac{1}{2m}(A\theta - Y)^T(A\theta - Y), \tag{5}$$

where $Y = [y_1, y_2, \cdots, y_m]^T$ is a column vector of $m \times 1$, which consists of actual measurements (sample outputs), and T denotes the transposition operator. Then, the partial derivative of $J(\theta)$ with respect to θ can be computed via

$$\frac{\partial}{\partial\theta}J(\theta) = \frac{\partial}{\partial\theta}\frac{1}{2m}(\theta^T A^T A\theta - \theta^T A^T Y - Y^T A\theta + Y^T Y) = \frac{1}{m}A^T(A\theta - Y). \tag{6}$$

Thus, the updated expression of θ can be rewritten as

$$\theta^{\text{updated}} = \theta - \alpha \cdot \frac{1}{m}A^T(A\theta - Y). \tag{7}$$

Figure 1a shows the schematic diagram of gradient descent. The minimum value of $J(\theta)$ can be obtained by iterating along the direction of gradient descent. First, it is worth mentioning that the value α represents the length of each step in the gradient direction during iterations. If the value α is too large, we cannot guarantee that the gradient will be decreased in each iteration, nor can we guarantee the convergence. If the value α is too small, it will lead to a painfully slow convergence and long calculation time, but the update value can achieve almost an optimal solution to the objective function, as it is unlikely that the stepping will miss any useful gradient position. Thus, this value α determines

the convergence speed of the iterations and whether the iterations can reach the optimal solution. Second, if the function $J(\theta)$ is convex, the iterative result starting from one initial value will be the optimal solution by a large probability; if $J(\theta)$ is non-convex, it is necessary to perform the gradient descent strategy multiple times, each with different initial values, to get rid of the local optimal solution, and then the solution with the smallest functional value should be selected from these iterative results. Generally, different initial values may lead to different minimum values of the objective function. Therefore, we should carefully choose the initial value of iterations, as well as the iteration step size.

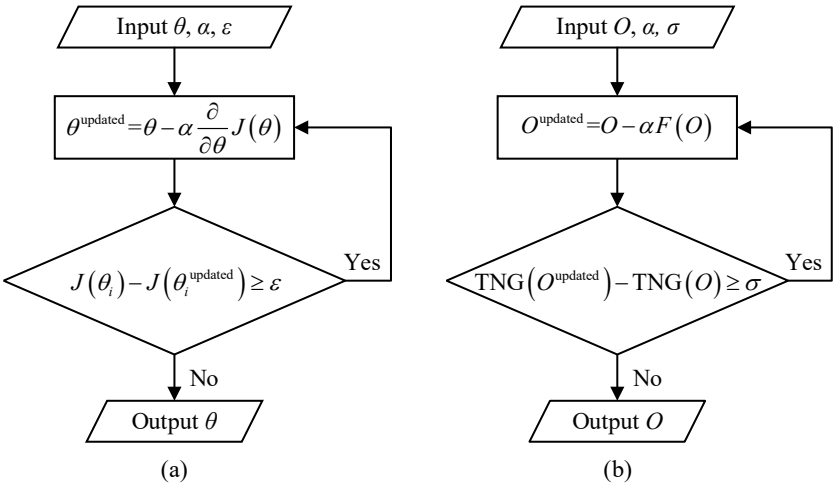


Figure 1. Algorithm flow chart. (**a**) Flow chart of the gradient descent idea. (**b**) Flow chart of our gradient-descent-like ghost imaging algorithm.

2.2. Intensity Correlation Functions in GI

In GI, the object $\widehat{O}(x,y)$ is illuminated by several modulated patterns $I_R(x,y) = \{I_R^1(x,y), I_R^2(x,y), \cdots, I_R^j(x,y), \cdots, I_R^m(x,y)\}$, where the superscript $j = 1,2,3,\cdots,m$ denotes the jth modulation, x and y stand for the spatial coordinates on x-axis and y-axis. The total intensity collected by the bucket detector can be written as $S_B = \int I_R(x,y)\widehat{O}(x,y)dxdy$. The object image can be retrieved by calculating the intensity correlation between the patterns $\{I_R^1(x,y), I_R^2(x,y), \cdots, I_R^m(x,y)\}$ and bucket values $S_B = \{S_B^1, S_B^2, \cdots, S_B^m\}$. In the following, we will briefly introduce BGI, HGI, DGI, logarithmic GI (LGI), trigonometric ghost imaging (TGI) [36].

First of all, the most classic second-order correlation in GI can be written as

$$G^{(2)}(x,y) = \langle S_B I_R(x,y)\rangle, \tag{8}$$

where $\langle u\rangle = \frac{1}{m}\sum_{j=1}^{m} u^j$ stands for the ensemble average of the signal u. It is worth mentioning that this classic second-order correlation function is a prototype algorithm of GI. Obviously, this function is too simple to obtain a satisfactory image quality, even with a large number of measurements, and the reconstructed image contains non-negligible background noise (which can be treated as the direct component). Therefore, this formula has basically withdrawn from history, but its derivatives have gradually become the mainstream algorithms.

For example, based on the above formula, by making S_B and $I_R(x,y)$ separately subtract the average terms of S_B and $I_R(x,y)$, we can acquire the functional form of the BGI [15]:

$$\begin{aligned} G_{\mathrm{BGI}}(x,y) &= \langle (S_B - \langle S_B\rangle)(I_R(x,y) - \langle I_R(x,y)\rangle)\rangle \\ &= \langle S_B I_R(x,y)\rangle - \langle S_B\rangle\langle I_R(x,y)\rangle. \end{aligned} \tag{9}$$

Thus, this formula uses $\langle S_B\rangle\langle I_R(x,y)\rangle$ to describe the background noise. By subtracting this product term from $\langle S_B I_R(x,y)\rangle$, the BGI can generate a good-quality image for simple

objects with high transmittance, but fails to work in the case of unstable light sources (e.g., with temperature drift) or quasitransparent complex objects.

By calculating high-order correlation [16], we can obtain

$$G_{\mathrm{HGI}}(x,y) = \left\langle (S_B)^p (I_R(x,y))^q \right\rangle, \tag{10}$$

where p and q are the power indices of the bucket and reference signals, respectively. The HGI can improve the visibility and contrast by selecting appropriate p and q values, but it cannot effectively remove background noise.

By replacing $\langle S_B \rangle$ in BGI with a new differential term $\frac{\langle S_B \rangle}{\langle S_R \rangle} S_R$ ($S_R = \int I_R(x,y)dxdy$), which reflects the relative fluctuations in the bucket values, we will obtain the functional form of the DGI [18]:

$$\begin{aligned} G_{\mathrm{DGI}}(x,y) &= \left\langle (S_B - \frac{\langle S_B \rangle}{\langle S_R \rangle} S_R)(I_R(x,y) - \langle I_R(x,y) \rangle) \right\rangle \\ &= \langle S_B I_R(x,y) \rangle - \frac{\langle S_B \rangle}{\langle S_R \rangle} \langle S_R I_R(x,y) \rangle. \end{aligned} \tag{11}$$

Compared with the BGI, the term $\frac{\langle S_B \rangle}{\langle S_R \rangle} \langle S_R I_R(x,y) \rangle$ in the DGI can better describe the background noise. Therefore, the DGI can more accurately remove background noise, and significantly improve the image quality, even in harsh or noisy measurement environments.

The expression of the LGI is defined [36] as follows:

$$G_{\mathrm{LGI}}(x,y) = \left\langle \log_C \frac{S_B}{\langle S_B \rangle} I_R(x,y) \right\rangle, \tag{12}$$

where C is the base of the logarithm function. According to the logarithmic operational rule, the above formula can also be rewritten as $G_{\mathrm{LGI}}(x,y) = \langle \log_C S_B I_R(x,y) \rangle - \langle \log_C \langle S_B \rangle \rangle$; thus, some part of background noise is removed.

The formula of the TGI [36] can be written as

$$G_{\mathrm{TGI}}(x,y) = \left\langle [\sin(d + S'_B - \frac{1}{2})\pi] I_R(x,y) \right\rangle, \tag{13}$$

where $S'_B = \left(S_B - S_B^{\min}\right) / \left(S_B^{\max} - S_B^{\min}\right)$, $S_B^{\max}$ and $S_B^{\min}$ are the maximum and minimum of S_B, respectively. Here, d takes even integers to generate positive images [36]. Since the TGI only performs a triangular transformation on S'_B on the basis of the classic second-order correlation function, the reconstruction quality obtained via TGI is not very different from the recovered result of the classic second-order correlation function.

We find that Equations (8)–(13) can be divided into two categories: functions without subtractive background terms and functions with subtractive background terms. The first category, containing the classic second-order correlation function, HGI and TGI, mainly focuses on enhancing the image visibility by optimizing bucket values, modulated patterns or both; the second category, including the BGI, DGI and LGI, tries to characterize the background noise and subtracts it from the reconstructed image of the classic second-order correlation function to improve the image quality. Among them, the DGI is recognized as the best statistical correlation algorithm due to its excellent and stable imaging performance.

To simplify the derivation, we will also derive the matrix forms of these functions. Each modulated pattern can be reshaped into a row vector of $1 \times n$, m row vectors reshaped from m patterns will form a measurement matrix A of $m \times n$. In the same way, the object image $\widehat{O}$ can be flattened into a column vector O of $n \times 1$. Thus, the ideal bucket values can be written as AO, and the actual bucket values will be denoted by $V = [S_B^1, S_B^2, \cdots, S_B^m]^T$. In the ideal measurement environment, $V = AO$. Then, we will have $\langle S_B \rangle = \frac{1}{m} V^T E_{m\times 1} = \frac{1}{m} E_{1\times m} V$, $\langle I_A \rangle = \frac{1}{m} A^T E_{m\times 1}$, $S_A = A E_{n\times 1}$, $\langle S_A \rangle = \frac{1}{m} E_{1\times m} A E_{n\times 1}$, where E denotes a matrix consisting of all ones and its subscripts stand for the dimensions of this matrix. In the following, we can rewrite the above intensity correlation functions as

$$F_{G^{(2)}}(O) = \frac{1}{m} A^T AO, \tag{14}$$

$$F_{G_{\mathrm{BGI}}}(O) = \frac{1}{m} A^T AO - \frac{1}{m^2} E_{1\times m} AO A^T E_{m\times 1}, \tag{15}$$

$$F_{G_{\mathrm{HGI}}}(O) = \frac{1}{m} (A^T)^{\cdot q} (AO)^{\cdot p}, \tag{16}$$

$$\begin{aligned} F_{G_{\mathrm{DGI}}}(O) &= \frac{1}{m} A^T AO - \frac{\frac{1}{m} E_{1\times m} V}{\frac{1}{m} E_{1\times m} AE_{n\times 1}} \cdot \frac{1}{m} A^T AE_{n\times 1} \\ &= \frac{1}{m} A^T A \left(O - \frac{E_{1\times m} AOE_{n\times 1}}{E_{1\times m} AE_{n\times 1}} \right) \\ &= \frac{1}{m} A^T A (O - \widetilde{O}) \\ &= \frac{1}{m} A^T (AO - A\widetilde{O}) \\ &= \frac{1}{m} A^T (AO - \widetilde{V}), \end{aligned} \tag{17}$$

$$F_{G_{\mathrm{LGI}}}(O) = \frac{1}{m} A^T \left(\log_C \frac{AO}{\frac{1}{m} E_{1\times m} AO} \right), \tag{18}$$

$$F_{G_{\mathrm{TGI}}}(O) = \frac{1}{m} A^T \left(\sin\left(d + V' - \frac{1}{2} \right) \pi \right), \tag{19}$$

where $.q$ and $.p$ denote the powers that are performed on each element of the matrix, $\widetilde{O} = \frac{E_{1\times m} AOE_{n\times 1}}{E_{1\times m} AE_{n\times 1}}$ characterizes the average light field, $\widetilde{V} = A\widetilde{O}$, and $V' = (V - V_{\min})/(V_{\max} - V_{\min})$. Although the above derivations are carried out under noise-free conditions, the actual recorded bucket values are generally accompanied by the scaling of light intensities as well as measurement noise e, i.e., $V = \kappa AO + e$, where κ denotes the proportional coefficient determined by the light attenuation, collection efficiency, photoelectric conversion efficiency, etc. However, this does not affect the above derivations because, for $V = \kappa AO + e = A(\kappa O + e') = AO'$, where $O' = \kappa O + e'$, we only need to replace original O with O'.

2.3. Gradient-Descent-like Ghost Imaging

Next, let us first review the composition of the original image O. Generally, the image O can be considered to be composed of the object part O_o and background part O_b (a direct component), i.e., $O = O_o + O_b$ (the values in both O_o and O_b are greater than 0); thus, we will have $AO = AO_o + AO_b$. According to the previous classification, it is easy to notice that $F_{G_{\mathrm{BGI}}}(O)$, $F_{G_{\mathrm{DGI}}}(O)$, $F_{G_{\mathrm{LGI}}}(O)$ are in forms of $\frac{1}{m} A^T (AO - \widetilde{\Omega})$. In terms of GI principle, we expect to make the element values of this subtractive term $\widetilde{\Omega}$ as close to those of AO_b as possible. The closer these two sets of values are, the higher the reconstruction quality of the intensity correlation function is. We find that the form of $\frac{1}{m} A^T (AO - \widetilde{\Omega})$ looks very similar to the gradient form $\frac{\partial}{\partial \theta} J(\theta) = \frac{1}{m} A^T (A\theta - Y)$ of the traditional gradient descent algorithm. If we regard the above three intensity correlation functions as the image gradients $\frac{\partial}{\partial O} J(O)$, then their primitive functions $J(O)$ can be written as the transform forms of $J(O) = \frac{1}{2m} (AO - \widetilde{\Omega})^T (AO - \widetilde{\Omega})$. $F_{G_{\mathrm{DGI}}}(O)$ happens to be the gradient of the loss function $J(O) = \frac{1}{2m} (AO - \widetilde{V})^T (AO - \widetilde{V})$ without any transformation. Different from the loss function (e.g., $J(\theta) = \frac{1}{2m} (A\theta - Y)^T (A\theta - Y)$) in traditional gradient descent algorithm, here, $J(O)$ is to make $\widetilde{\Omega}$ as close to a direct component AO_b (rather than AO) as possible so that $J(O)$ will reach its minimum near the optimal solution (while in classic gradient descent algorithm, $J(\theta)$ is to make Y as close to $A\theta$ as possible so that the residual error can

be minimized; thus, both $J(\theta)$ and its gradient will vanish to 0 near the optimal solution). However, it does not affect the correlation functions being regarded here as the gradients, because, no matter whether $J(\theta)$ or $J(O)$ is used, the ultimate goal is to minimize the objective function by reducing the gradient value. Following this idea, $A^T\widetilde{\Omega} \to A^T A O_b$ tries to characterize the background noise, so $A^T AO - A^T\widetilde{\Omega}$ intends to extract the object part, and the values of $\frac{1}{m}A^T(AO - \widetilde{\Omega})$ are expected to be continuously reduced to make the reconstructed object part closer to the optimal solution. $F_{G_{\mathrm{HGI}}}(O)$ and $F_{G_{\mathrm{TGI}}}(O)$ can be expressed as the transform forms of $F_{G^{(2)}}(O)$ ($\frac{1}{m}A^T AO$), namely, to make the subtractive term $\widetilde{\Omega}$ in $\frac{1}{m}A^T(AO - \widetilde{\Omega})$ equal to zero, but with relatively poor imaging performance. Therefore, most of the intensity correlation functions (especially the ones with subtractive background terms) can be regarded as the gradients that are expected to be minimized.

According to the above theoretical analysis and proof, we propose a GGI method. It has been proven that, no matter what correlation function we use to recover the object image, the probability of recovered pixel values located in the pixel region of the same original gray value of the object follows a Gaussian distribution, whose average value has a linear relationship with this given original gray value [31,37,38]. For this reason, all intensity correlation algorithms can be regarded as the functions F of the object image $\widehat{O}$, denoted as $F(\widehat{O})$. Here, we let $F(O)$ (with respect to the current estimate image O) represent the above intensity correlation functions. $F(O)$ can be regarded as the gradient of $J(O)$, i.e., the partial derivative $\frac{\partial}{\partial O}J(O)$. Since the original object image to be recovered is generally unknown in advance, $J(O)$ will fail to work because its change cannot be observed intuitively. Fortunately, there are many equivalent loss functions available as alternatives to assert the convergence of the iterations in a direct way, such as $\|O_k - O_{k-1}\|_0$, $\|O_k - O_{k-1}\|_1$, $\|O_k - O_{k-1}\|_2$, $\|O_k - O_{k-1}\|_{\mathrm{TV}}$ and no-reference image quality assessment, where TV is short for total variation. Here, we choose to use Tenengrad (TNG) function [39,40] (a similar method to the TV norm) as the algorithm's termination criterion, which calculates the sum of the squares of all pixels' gradient values in a current estimate that are greater than a certain threshold, and is defined [39] as

$$\mathrm{TNG} = \sum_x \sum_y [G(x,y)]^2, \quad \text{for } G(x,y) > \gamma, \tag{20}$$

where γ is the preset threshold. In this function, $G(x,y)$ is the gradient, which can be written as

$$G(x,y) = \sqrt{\left(f_x^2 + f_y^2\right)}, \tag{21}$$

where f_x and f_y are the gradients along x and y axes, respectively, which can be obtained by the well-known Sobel (discrete differentiation) operator. Generally, the sharper the edge of the natural image, the higher the gradient value (before computing the value of TNG, the column vector O needs to be reshaped into a matrix form). In its implementation, we only need to compare the difference between the values of the TNG function obtained by two adjacent iterations to determine whether the algorithm has converged.

Figure 1b gives the algorithm flow chart of the proposed GGI, and its algorithm steps can be described as follows:

(1) Compute the current $F(O)$ with respect to the current estimate O;
(2) Let O minus α times the current estimate obtained in Step (1) to produce an updated estimate O^{updated}:

$$O^{\mathrm{updated}} = O - \alpha F(O); \tag{22}$$

(3) Judge whether the difference between $\mathrm{TNG}\left(O^{\mathrm{updated}}\right)$ and $\mathrm{TNG}(O)$ is greater than a predefined threshold σ (which is set to 10^{-10} in this work): if not, the algorithm will be terminated, the current estimate O is the optimal solution, which will then be reshaped to an image matrix U; if so, replace O with O^{updated} and repeat Steps (1–2);

(4) Output the reconstructed image U.

It should be noted that the main iterative formula as given in Equation (22) is only related to the current estimate O (a possible solution to the recovered ghost image), predefined threshold α and newly computed function $F(O)$, while intensity correlation functions depend on the current bucket values and modulated patterns. This means that the function $F(O)$ in Equation (22) can be replaced with intensity correlation functions, with the inputs serving as the modulated patterns and the new bucket values calculated with the current estimate O. In other words, in each iteration, the modulated patterns remain unchanged, while the bucket values are updated according to the current estimate O. Therefore, the bucket values used in the first iteration are obtained from actual measurements, while the ones used in the subsequent iterations are numerically calculated. In addition, since the intensity correlation function $F(O)$, loss function $J(O)$ and its equivalent TNG function are all convex, by following the direction of the slope of the convex surface created by the loss function downhill, the iterations that start from some initial values tend to reach a valley, and the (local) optimal solution will be acquired after multiple iterations by a very large probability according to the theory of gradient descent. The iterative formulas that use intensity correlation functions with subtractive background terms will definitely achieve better convergence than the cases using intensity correlation functions without subtractive background terms, because the gradient functions in the former cases will tend towards the minimum near the optimal solution and are more conducive to the convergence of iterations.

Next, for the sake of simplification, we abbreviate the gradient-descent-like BGI, gradient-descent-like HGI, gradient-descent-like DGI, gradient-descent-like LGI and gradient-descent-like TGI as the GBGI, GHGI, GDGI, GLGI, GTGI, respectively.

3. Simulation Results and Analysis

To demonstrate the performance of the proposed method, we used random binary patterns for numerical simulations. Here, we set $p = 10$ and $q = 1$ for the HGI function, as a large p will increase the image visibility and a small q will dramatically suppress the noise. For LGI, the base of the logarithmic function is set to $C = 10$, while, in TGI, we set $d = 0$.

Figure 2a shows the binary image labelled "01" consisting of 64×64 pixels. The DGI, GBGI, GHGI, GDGI, GLGI and GTGI are all computed with 20,480 measurements and the step size α is set at 0.1 (which will be explained below). It should be noted that all these reconstruction algorithms are based on statistical intensity correlation, which means that oversampling is generally required in classic GI functions, but the large-scale matrix multiplication and prior knowledge of image sparsity are no longer needed, so the computational complexity is very low. Although oversampling is not as efficient as a simple scanning, GI, as an indirect computational imaging technique, only needs to record the total light intensity, which can effectively avoid the average distribution of luminous flux across dimensions, as in point scanning. Especially under ultra-low-light illumination, this advantage will become more prominent. In addition, GI has a better anti-turbulence and a better imaging sensitivity than scanning. Since the focus of this manuscript is more on how to improve the imaging quality of the intensity correlation functions through only a few iterations, the number of measurements used here (20,480) is just an example to illustrate the performance improvement, and the subsampling cases will be discussed later. The reconstruction results of the above algorithms are given in Figure 2b–g, with their SNRs and running time t marked right below the figures. The algorithms are running under the operating system Windows 10, with the memory of 16.0 GB and the central processor AMD Ryzen 7 4800H of 2.90 GHz integrated with Radeon Graphics. It can clearly be seen from the simulation results that the SNRs of the GBGI, GDGI, and GLGI are higher than those of the GHGI and GTGI (consistent with our previous theoretical analysis), which means that they gradually approach the position of the optimal solution after a few iterations. We note that the SNR of the GHGI is slightly lower than that of DGI, because the high-

order operation of the bucket values amplifies the noise while enhancing the signal. The GTGI also fails to acquire the performance gain, mainly due to the absence of subtractive background term. It is worth noting that, among these, the SNR of the GDGI can reach 9.60, which is several times that of the DGI. The main reason the GDGI can outperform other variants is that the DGI function expression it uses happens to be $\frac{1}{m}A^T(AO - \widetilde{V})$, whose primitive function is exactly the loss function $J(O) = \frac{1}{2m}(AO - \widetilde{V})^T(AO - \widetilde{V})$ without any transformation. Besides this, the calculation time of the GDGI is 3.57 s, with only a small increase in computing time compared to the DGI, which is acceptable in most cases. Therefore, the GDGI can achieve a much better reconstruction performance without much increase in computing time compared with the DGI.

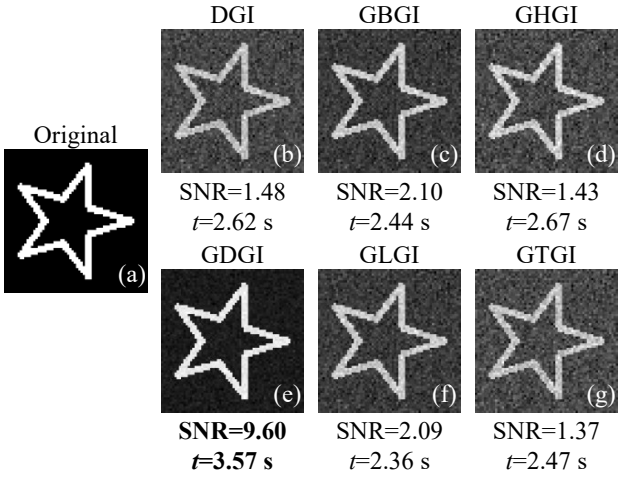


Figure 2. Simulation results of the binary image "01" using different ghost imaging (GI) algorithms with $m = 20,480$ measurements. (**a**) is the original image "01", (**b**–**g**) are the images recovered by the differential GI (DGI), gradient-descent-like background-removal GI (GBGI), gradient-descent-like high-order GI (GHGI), gradient-descent-like DGI (GDGI), gradient-descent-like logarithmic GI (GLGI) and gradient-descent-like trigonometric GI (GTGI), respectively, with their signal-to-noise ratios (SNRs) and running time t marked below the figures.

Figure 3a shows a 64×64 grayscale image labelled "02", whose gray value range is [0, 1]. With the same number of measurements, the above algorithms are executed again. Here, we set $\alpha = 0.1$. The results are presented in Figure 3b–g. The SNR and calculation time of the DGI is 3.86 and 2.59 s, respectively. In the recovered image of the DGI, the contour of the object is recognizable, but the background noise is messy and the detailed information is missing. As we can see, the GDGI acquires the best performance, with an SNR value of up to 32.70. The image retrieved by the GDGI contains only a small amount of noise and shows a great ability to resolve the details. In addition, the SNRs of the GBGI and GLGI are more than three times that of the DGI, while the GHGI's performance was the worst (as it hardly acquires any useful object information) among these GGI variants due to the aforementioned HGI restrictions. The calculation time of the above methods is very close, with all less than 4 s, which further indicates that the GDGI can obtain high-quality reconstructed images, with the noise being well-suppressed at the expense of a negligible increase in computing time.

As mentioned above, the gradient descent results are greatly influenced by the step size α. Given that the GDGI has the best performance among the above variant algorithms, without loss of generality, we will mainly focus on the GDGI algorithm and discuss the relationship between the SNR and α value in depth. Take the binary image "01" and grayscale image "02" ranging from 0 to 1, as shown in Figures 2a and 3a for examples, we drew the SNR curves of the GDGI as functions of the step size α as well as the calculation time t for the original images "01" and "02", respectively, as shown in Figure 4a–d. The

step size α was set from 0.005 to 0.5 with a 0.005 stepping increase, and, for each step size, both the SNR and calculation time were recorded.

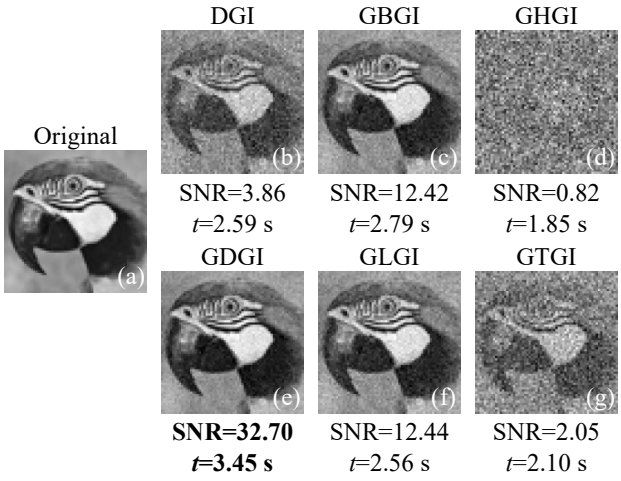


Figure 3. Simulation results of the grayscale image labelled "02" using different algorithms with $m = 20,480$ measurements. (**a**) is the original image "02", (**b**–**g**) are the images retrieved by the DGI, GBGI, GHGI, GDGI, GLGI and GTGI, respectively.

As can be seen from Figure 4a,b, with the increase in the step size α, the SNRs of the reconstructed images of "01" present an oscillating downward trend, and the calculation time t decreases sharply and tends to be stable when $\alpha \geq 0.2$. As we know, when the step size α is small, the gradient descend stepping distance of the each iteration is too small, so the convergence speed of iterations is slow and the calculation time is long. However, in this case, the algorithm is unlikely to miss the optimal solution in its iterative process, leading to a relatively high SNR with a high probability. When the step size α becomes large, the iterative convergence speed increases rapidly, but the iteration precision will drop, and there is a high probability that the optimal solution will be missed, causing a relatively low SNR. Furthermore, for some large α values, the current step of the iterations will cross the position of the optimal solution, which will lead to fluctuations, as shown in Figure 4a,c. Taking the DGI results of the binary image "01" (i.e., SNR = 1.48 and t = 2.62 s, see Figure 2b as a reference (which is drawn as straight lines in Figure 4a,b), it can be seen from the curves that no matter how the value of α changes in the range from 0 to 0.5, the SNRs of the GDGI are always higher than those of the DGI, and when α is greater than 0.1, the calculation time of GDGI is almost the same as that of the DGI. The above results demonstrate that our proposed method has a more efficient imaging performance than the DGI in the binary case.

For the grayscale image "02", from Figure 4c,d we can see that the SNRs of reconstructed images present an overall downward trend and the calculation time t also drops sharply with the increase in step size α. When $0 < \alpha \leq 0.1$, the SNRs of the GDGI are almost constant. When α further increases, the SNR gradually changes from a small-amplitude fluctuation to a large-amplitude fluctuation, as shown in Figure 4c. This is because the increase in the step size α causes an increase in the gradient descent stepping distance (when the iteration is close to the optimal solution but the preset termination conditions are not met, the jump will continue; the larger the step size, the more likely it is to jump out of the region where the optimal solution is located and enter a new reciprocating iteration, which may cause reciprocating oscillations in SNR values). Similarly, we took the DGI results of the grayscale image "02" (i.e., SNR = 3.86 and t = 2.59 s, see Figure 3b) as a reference (which is also drawn as straight lines in Figure 4c,d). We can see from Figure 4c,d that the SNRs of the GDGI never fall below those of the DGI, and when α is greater than 0.1, the calculation times of the GDGI and DGI are nearly the same. This proves that our scheme is also suitable for complex grayscale objects. According to the above results, to

form a trade-off between the imaging quality and running time, the appropriate value range of the step size α should range from 0.05 to 0.15. For simplicity, in the following, we set $\alpha = 0.1$. In addition, since the result of $F(O)$ is always positive, the iterative updating direction is usually opposite to the gradient direction, but we believe that α, in some special cases, can also take a negative value; for example, when the step size is too large, the algorithm misses the optimal solution and α should be changed to a negative value. This appears complicated and needs further study in the future, so it will not be discussed in this article.

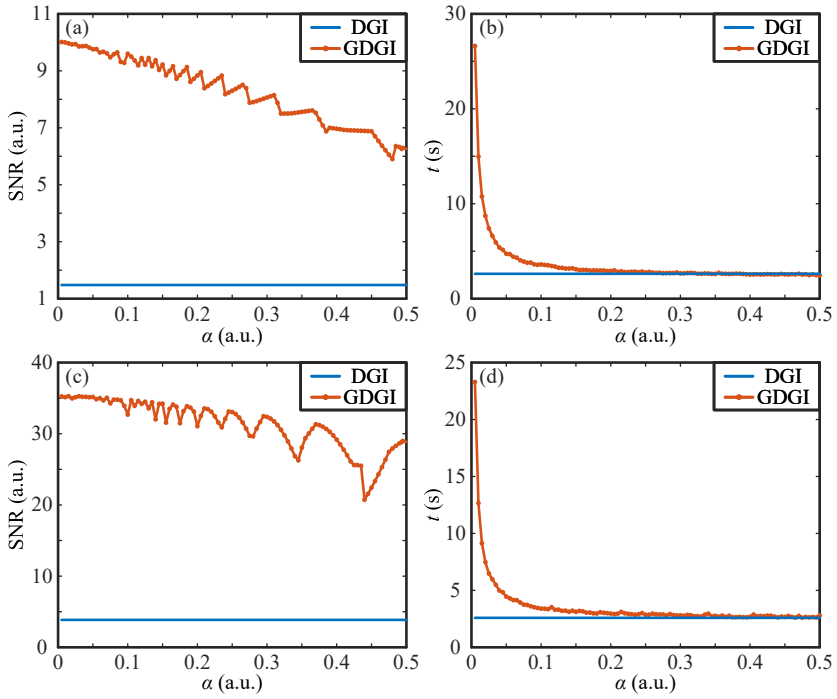


Figure 4. SNR and t curves as functions of the step size α. (**a**,**b**) and (**c**,**d**) are the curves for objects "01" and "02" by using the DGI and GDGI, respectively.

Next, to verify the universality and effectiveness of our algorithm, we tested another binary image and a grayscale image whose gray value range is [0, 1], marked as "03" and "04", respectively. Figure 5 shows the original images and the corresponding imaging results of the DGI, PGI and GDGI with different sampling rates r (defined as the ratio of the number of measurements to the total number of image pixels). As can be seen from Figure 5b–m,o–z, when r = 0.5, the image quality of the PGI is slightly better than those of the DGI and GDGI for the image "03", while the image quality of the GDGI is 0.59 and 0.82 times higher than those of the DGI and PGI for the image "04". When r = 1, the image quality of the GDGI is 0.08 and 0.14 times higher than those of the DGI and PGI for the image "03", and 1.99 and 2.57 times higher than those of the DGI and PGI for the image "04". When r = 3 and 5, the imaging performance of the GDGI is much better than those of the DGI and PGI. Therefore, it can be inferred that, with the increase in sampling rate r, the improvement in image quality will be more significant. In addition, since our algorithm does not require preprocessing or optimizing of the measurement data, the use of a gradient-descent-like method can also fully improve the utilization of data, thereby greatly saving additional computational overhead. Here, the average running time of the GDGI is only 1.19 and 1.18 times those of the DGI or PGI, respectively, which means that a significant improvement in imaging quality can be achieved without much running time increment, i.e., our scheme has a relatively high return-to-investment ratio. Besides, through this simulation, it also demonstrates that our algorithm works well for different types of images.

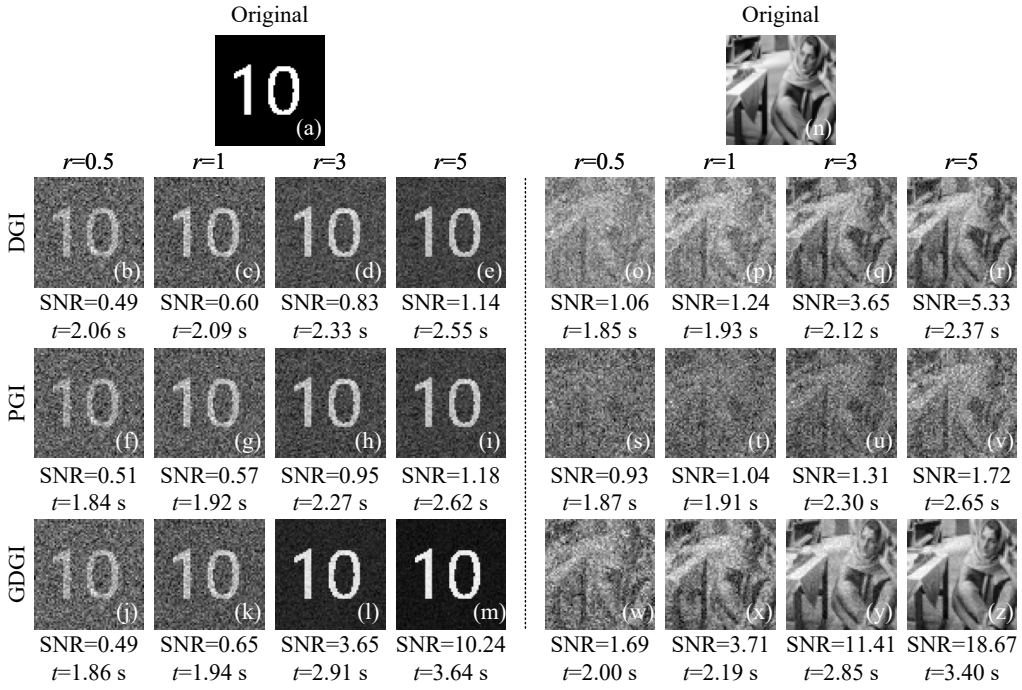


Figure 5. Simulation results of another two images "03" and "04" (all of 64 × 64 pixels) using the DGI, pseudo-inverse GI (PGI) and GDGI algorithms with $r = 0.5, 1, 3, 5$ sampling rates. (**a**,**n**) are the original images, (**b**–**e**) and (**o**–**r**), (**f**–**i**) and (**s**–**v**), (**j**–**m**) and (**w**–**z**) are the images recovered by the DGI, PGI and GDGI, respectively.

In the following, we performed additional numerical simulations to test the robustness of the GDGI against noise. Taking the image "02" as an example, we further added white Gaussian noise to the bucket value S_B to simulate the noisy measurements. Then, we calculated the SNRs of reconstructed images under different detection signal-to-noise ratios (DSNRs), i.e., the power ratio of the signal to the measurement noise, which can be expressed as $\mathrm{DSNR} = 10\log_{10}(Var(S_B)/Var(S_{noise}))$ [29], where $Var(S_B)$ and $Var(S_{noise})$ denote the variances of the bucket values and measurement noise, respectively. As shown in Figure 6a, when the Gaussian random noise with different variances is added to the bucket signal, the SNRs of images recovered by the DGI, PGI and GDGI decrease with the decrease in the DSNR. It can be clearly seen from the curves that when the DSNR is greater than 5 dB, the imaging performance of the GDGI is significantly better than those of the DGI and PGI; when the DSNR is less than 5 dB, the SNRs of these three methods become closer, but the SNRs of the GDGI are still slightly higher than those of the DGI and PGI. This is because the statistical averaging of the intensity correlation is accompanied by reconstruction noise. The lower the DSNR, the harsher the measurement condition and the more difficult it is to distinguish the useful signal from the noise, which leads to a larger reconstruction noise and makes it more difficult to recognize the object part in the image. In ultra-low DSNR conditions, all algorithms including convex optimization will tend to fail. As shown in Figure 6b–j, we made some comparisons among the DGI, PGI and GDGI, with the DSNR changing from 35 dB to 15 dB. When DSNR = 35 dB, 25 dB and 15 dB, the SNRs of the GDGI are 8.87, 6.71 and 3.13 times those of the DGI, and are 16.91, 12.79, 5.87 times those of the PGI, showing that our algorithm has a good anti-noise ability compared with the traditional GI algorithm under noisy measurements.

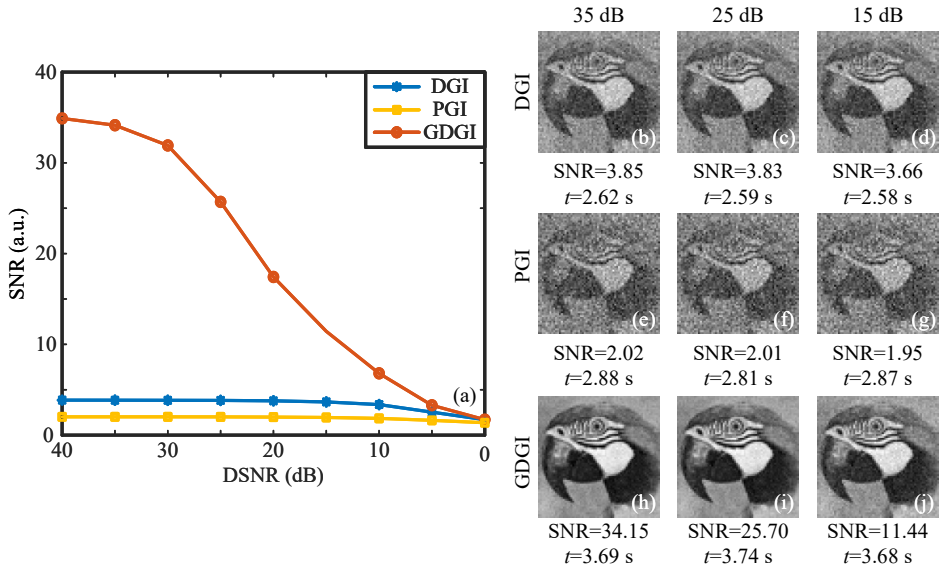


Figure 6. Simulation results for the grayscale image labelled "02" using the DGI, PGI and GDGI algorithms under different detection signal-to-noise ratios (DSNRs). (**a**) gives the changes in the SNRs of the DGI, PGI and GDGI with the DSNRs, (**b**–**d**), (**e**–**g**) and (**h**–**j**) are the images retrieved by the DGI, PGI and GDGI, with the DSNR being 35 dB, 25 dB, 15 dB.

4. Experimental Demonstration and Discussions

4.1. Experimental Setup

To further verify the effectiveness of the proposed architecture, we performed an optical experiment based on a classic single-arm computational GI setup, as shown in Figure 7, where a digital micromirror device (DMD) was used as an SLM and a counter-type Hamamatsu H10682-210 photomultiplier tube (PMT) was used as a bucket detector. The thermal light emitted from a halogen lamp passed through a collimator, an aperture diaphragm and a neutral density filter. Then the light beam illuminated the first DMD (indicated as DMD_1) and was modulated by the preset binary patterns. After that, the modulated (structured) light was projected onto the second DMD (indicated as DMD_2) through the imaging lens, where the digital object was displayed onto DMD_2. The light reflected by DMD_2 was converged by the lens onto the bucket detector. The detector recorded the sequence of the total intensity in the form of photon counts, denoted as $S_B = \{S_B^1, S_B^2, \cdots, S_B^j, \cdots, S_B^m\}$ (corresponding to different modulated binary patterns). In the following experiments, DGI and GDGI are the two main algorithms to be considered and compared.

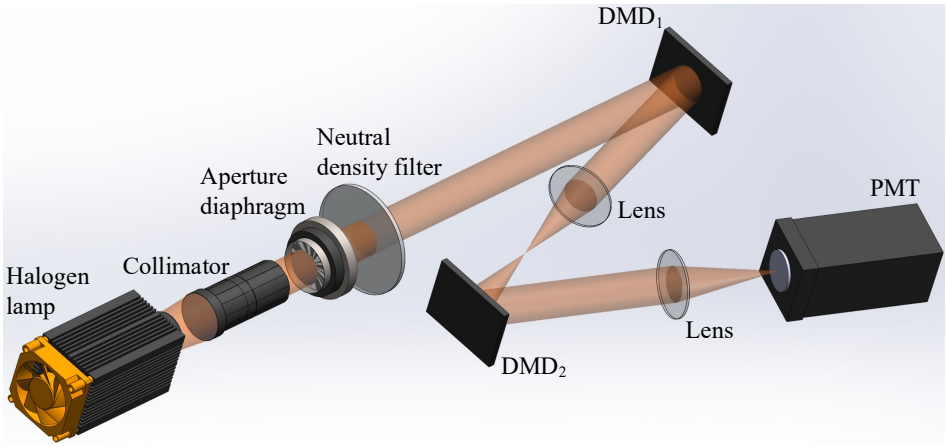


Figure 7. Experimental setup of computational GI. DMD: digital micromirror device, PMT: photomultiplier tube.

4.2. Experimental Results of a Binary Object

For the imaging experiments of a binary object, we encoded a 64 × 64 binary image "BIT" onto DMD_2 as the original object image (see Figure 8a), then presented the reconstructed images with $r = 5$ (as an example) in Figure 8b,c. The SNRs of the DGI and GDGI are 0.78 and 2.79, respectively. It is obvious that the latter is 3.58 times the former, but the calculation time of the latter is only 3.77 s, consistent with the aforementioned simulation results. Furthermore, the visibility of the GDGI is much better than that of the DGI. To make a more intuitive comparison of the reconstructed images, we plotted the curves of the grayscale values in the 32th row (consisting of 64 pixels) of the original image and restored images (marked by the red lines in the Figure 8a–c. It can be clearly seen from Figure 8d that our GDGI can significantly reduce the background noise and improve the contrast compared with traditional DGI, thus dramatically improving the visibility. Without loss of generality, this also indicates that our GGI scheme can efficiently improve the SNRs of GI reconstructions.

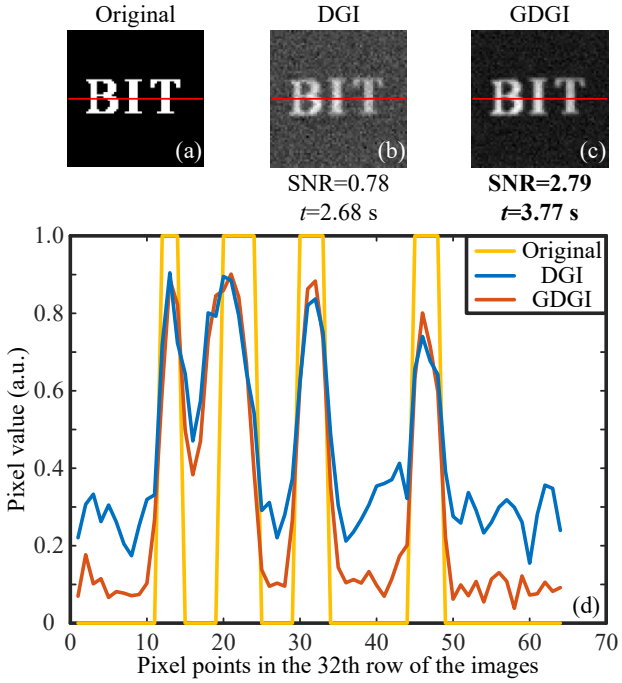


Figure 8. Experimental results of a binary object. (**a**) is the original image encoded onto DMD_2, (**b**,**c**) are the experimental results recovered by the DGI and GDGI, and (**d**) gives cross-section plots of the grayscale images in the 32th row.

Next, the images of the DGI and GDGI were reconstruced using different sampling rates (see the left half of Figure 9). Above the dotted line on the left half of the graph, we increased the sampling rate r from 0.25 to 1.25, step by 0.25, and below the dotted line, we made r change from 4 to 5, also step by 0.25. It is easy to see from Figure 9a–t that when r is smaller than 1, the GGI benefits are small, and when r is larger than 1, the gains become increasingly apparent. The reason for this is that when the measurement data are relatively limited, it is difficult for a finite number of iterations to maximize the performance gains; when the measurement data are rich, these iterations can effectively take advantage of the correlation within the data to suppress noise and improve image quality. To further demonstrate the performance of the proposed method, the relationship between the number of iterations it takes for the algorithm to terminate and the sampling rate is given on the right half of Figure 9. As shown in Figure 9u, when r is between 0.25 and 1.25, most of the iteration numbers required for the algorithm to converge are lower than 8. In this case, the GGI fails to obtain effective performance gains. However, when r is between 4∼5, the number of iterations required for convergence is about 16, as

shown in Figure 9v. In this situation, the iterations gradually approach the position of the optimal solution, and thus the image quality is significantly improved. Therefore, it can be concluded that, in our algorithm, the image quality can be effectively improved with a limited number of iterations without any preprocessing or optimization.

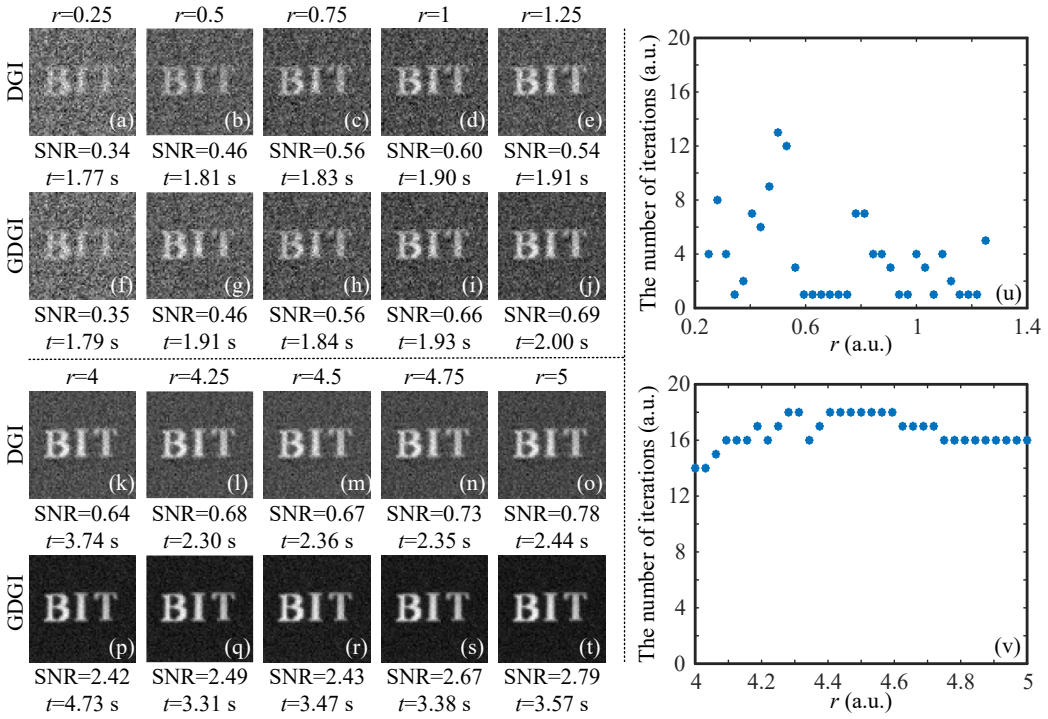


Figure 9. Comparison results of the binary object "BIT". (**a**–**t**) are the reconstructed images of the DGI and GDGI using different sampling rates; (**u**,**v**) give the change in the number of iterations with the sampling rate, all using the GDGI.

To more clearly see the change in the SNR values, we plotted the SNR curve as a function of the sampling rate r, as shown in Figure 10 (here, we took the sampled data corresponding to the original image "BIT" for investigation). One can see clearly that, with the increase in the sampling rate r, the SNRs of the GDGI increase more sharply than those of the DGI. Therefore, our nonlinear algorithm has a better performance than the linear algorithm.

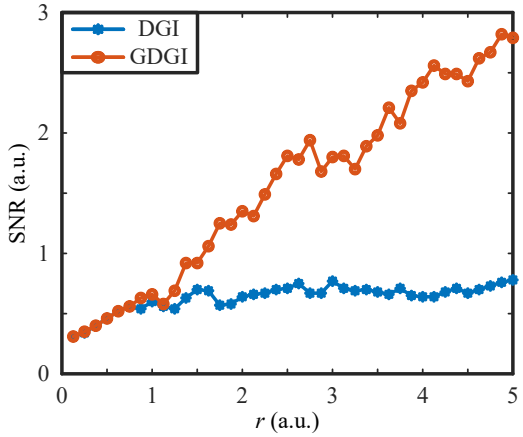


Figure 10. SNR values of the images reconstructed by the DGI and GDGI for the binary object "BIT" with different sampling rates r.

4.3. Experimental Results of a Grayscale Object

In the experiment, we use the image displayed on DMD_2 as the object. The DMD is composed of millions of micromirrors (pixels), each of which orientates $\pm 12^\circ$ with respect to the normal direction of the working plane, corresponding to two states (1 or 0). Thus, the matrix encoded onto the DMD should be binary, consisting of 1 and 0. To our knowledge, to display grayscale objects, the pulse-width modulation (PWM) strategy, which exchanges time-pulse width for grayscale level, is generally used. Since the DMD should sequentially switch multiple binary patterns for each grayscale image's display, the higher the grayscale level, the more time the PWM takes. Here, we chose to apply another method called Floyd–Steinberg error diffusion dithering [41] to convert the grayscale image to binary image. As shown in Figure 11, to reconstruct a grayscale image of 64×64, we need to apply upsampling to this image and set its each pixel as a pixel-unit "pix" consisting of 12×12 pixels (i.e., a grayscale pixel-unit occupies 144 mircormirrors in total with a gray value ranging from 0 to 144). The pixels are randomly lit up (set to 1) in a pixel-unit; the number of bright (1) pixels indicates the gray value of this pixel-unit. Here, for a 64×64 grayscale image's display, a matrix of 768×768 actual pixels will be loaded onto DMD_2.

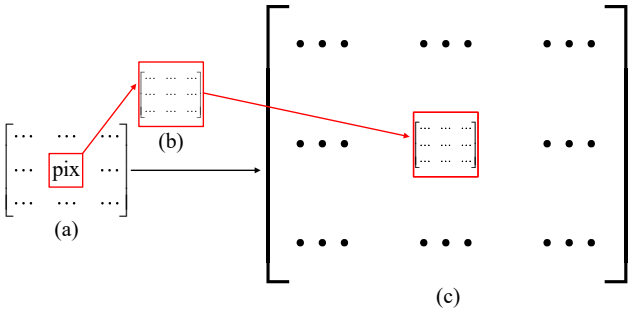


Figure 11. Schematic diagram of encoding a grayscale image onto a DMD. (**a**) shows the pixel size for reconstruction, (**b**) gives an example of an pixel-unit "pix", and (**c**) is the actual pixel size displayed on the DMD.

In this experiment, we encoded two grayscale images onto DMD_2 "car" (see Figure 12a,h) with a grayscale range of [0, 50] and [0, 100], respectively. Their experimental results using the DGI and GDGI under the sampling rate r = 5 and 10 were given in Figure 12b–e,i–l. It is obvious that the reconstruction results of r = 10 are better than those of r = 5, and the reconstructed details for a larger grayscale range are richer. To compare the details of the reconstructed images more intuitively, we enlarged the image regions within the red rectangles of Figure 12d–e,k–l, and drew green rectangles in the corresponding areas for more detailed observation, as shown in Figure 12f–g,m–n. Through subtle comparisons, we found that the GDGI will achieve a higher visibility than the DGI, but the improvement effects are not obvious. In this experiment, the measured bucket (single-pixel) signal is accompanied by a lot of noise, the DSNR is very low, and the GDGI algorithm terminates after only 1∼6 iterations; thus, the gain of the GDGI is small. The acquisition of high-quality images under low DSNR conditions is a tough problem for GI, and this will be the focus of our future research.

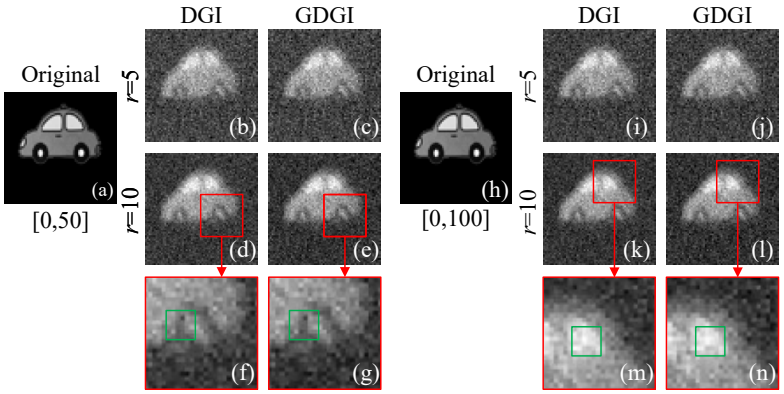


Figure 12. Experimental results of a grayscale object. (**a**,**h**) are the original object images "car" that are encoded onto DMD_2, with grayscale ranges of [0, 50] and [0, 100], respectively; (**b**–**e**) and (**i**–**l**) are the experimental results recovered by the DGI and GDGI, respectively; (**f**–**g**) and (**m**–**n**) are the enlarged detailed images of (**d**–**e**) and (**k**–**l**), respectively.

5. Conclusions

In conclusion, we present an effective GI reconstruction method based on gradient-descent-like strategy, which can use an intensity correlation function in its iterative formula. Both numerical simulation and optical experiments have been performed to demonstrate the performance of this proposed method. By setting the appropriate step size, this method can easily acquire an optimal solution through only a few iterations with low computational complexity. Moreover, it can also deal with both binary and grayscale complex objects in a slightly noisy environment, and suppress reconstruction noise to a certain extent. We find that the GDGI, as a variant of our GGI method, outperforms the other variants and can achieve a much better performance than that of the DGI, especially for binary complex objects, at the expense of a negligible increase in running time caused by iterations. Therefore, we believe that this new iterative calculation scheme may pave the way for the intensity correlation from the linear statistical average towards the nonlinear one, which will provide more possibilities for GI reconstructions.

Author Contributions: Conceptualization, W.-K.Y. and C.-X.Z.; methodology, W.-K.Y. and C.-X.Z.; software, W.-K.Y. and C.-X.Z.; validation, C.-X.Z.; formal analysis, W.-K.Y. and C.-X.Z.; investigation, W.-K.Y. and C.-X.Z.; resources, W.-K.Y.; data curation, C.-X.Z., Y.-X.L., S.-F.W. and C.C.; writing—original draft preparation, C.-X.Z.; writing—review and editing, W.-K.Y. and C.-X.Z.; visualization, C.-X.Z.; supervision, W.-K.Y.; project administration, W.-K.Y. funding acquisition, W.-K.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (61801022), the National Key Research and Development Program of China (2016YFE0131500), and the Civil Space Project of China (D040301).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Pittman, T.B.; Shih, Y.H.; Strekalov, D.V.; Sergienko, A.V. Optical imaging by means of two-photon quantum entanglement. *Phys. Rev. A* **1995**, *52*, R3429. [CrossRef] [PubMed]
2. Bennink, R.S.; Bentley, S.J.; Boyd, R.W. "Two-photon" coincidence imaging with a classical source. *Phys. Rev. Lett.* **2002**, *89*, 113601. [CrossRef] [PubMed]
3. Valencia, A.; Scarcelli, G.; D'Angelo, M.; Shih, Y. Two-photon imaging with thermal light. *Phys. Rev. Lett.* **2005**, *94*, 063601. [CrossRef] [PubMed]
4. Zhang, D.; Zhai, Y.H.; Wu, L.A.; Chen, X.H. Correlated two-photon imaging with true thermal light. *Opt. Lett.* **2005**, *30*, 2354–2356. [CrossRef] [PubMed]

5. Shapiro, J.H. Computational ghost imaging. *Phys. Rev. A* **2008**, *78*, 061802. [CrossRef]
6. Bromberg, Y.; Katz, O.; Silberberg, Y. Ghost imaging with a single detector. *Phys. Rev. A* **2009**, *79*, 053840. [CrossRef]
7. Bennink, R.S.; Bentley, S.J.; Boyd, R.W. Quantum and classical coincidence imaging. *Phys. Rev. Lett.* **2004**, *92*, 033601. [CrossRef]
8. Ferri, F.; Magatti, D.; Gatti, A.; Bache, M.; Brambilla, E.; Lugiato, L.A. High-resolution ghost image and ghost diffraction experiments with thermal light. *Phys. Rev. Lett.* **2005**, *94*, 183602. [CrossRef]
9. Xiong, J.; Cao, D.Z.; Huang, F.; Li, H.G.; Sun, X.J.; Wang, K. Experimental observation of classical sub-wavelength interference with thermal-like light. *Phys. Rev. Lett.* **2005**, *94*, 173601. [CrossRef]
10. Zhang, A.-X.; He, Y.-H.; Wu, L.-A.; Chen, L.-M.; Wang, B.-B. Table-top x-ray ghost imaging with ultra-low radiation. *Optica* **2017**, *5*, 374–377. [CrossRef]
11. Edgar, M.P.; Gibson, G.M.; Padgett, M.J. Principles and prospects for single-pixel imaging. *Nat. Photonics* **2018**, *13*, 13–20. [CrossRef]
12. Yu, W.-K.; Yao, X.-R.; Liu, X.-F.; Lan, R.-M.; Wu, L.-A.; Zhai, G.J.; Zhao, Q. Compressive microscopic imaging with "positive-negative" light modulation. *Opt. Commun.* **2016**, *371*, 105–111. [CrossRef]
13. Clemente, P.; Durán, V.; Torres-Company, V.; Tajahuerce, E.; Lanci, J. Optical encryption based on computational ghost imaging. *Opt. Lett.* **2010**, *35*, 2391–2393. [CrossRef]
14. Yu, W.-K. Cryptographic key distribution over a public network via variance-based watermarking in compressive measurements. *Appl. Opt.* **2019**, *58*, 5294–5300. [CrossRef] [PubMed]
15. Gatti, A.; Brambilla, E.; Bache, M.; Lugiato, L.A. Ghost imaging with thermal light: comparing entanglement and classical correlation. *Phys. Rev. Lett.* **2004**, *93*, 093602. [CrossRef] [PubMed]
16. Chan, K.W.C.; O'Sullivan, M.N.; Boyd, R.W. High-order thermal ghost imaging. *Opt. Lett.* **2009**, *34*, 3343–3345. [CrossRef]
17. Chan, K.W.C.; O'Sullivan, M.N.; Boyd, R.W. Optimization of thermal ghost imaging: high-order correlations vs. background subtraction. *Opt. Express* **2010**, *18*, 5562–5573. [CrossRef]
18. Ferri, F.; Magatti, D.; Lugiato, L.A.; Gatti, A. Differential ghost imaging. *Phys. Rev. Lett.* **2010**, *104*, 253603. [CrossRef]
19. Gong, W.-L. High-resolution pseudo-inverse ghost imaging. *Photonics Res.* **2015**, *3*, 234–237. [CrossRef]
20. Duarte, M.F.; Davenport, M.A.; Takhar, D.; Laska, J.N.; Sun, T.; Kelly, K.F.; Baraniuk, R.G. Single-pixel imaging via compressive sampling. *IEEE Signal Proc. Mag.* **2008**, *25*, 83–91. [CrossRef]
21. Katz, O.; Bromberg, Y.; Silberberg, Y.; Compressive ghost imaging. *Appl. Phys. Lett.* **2009**, *95*, 131110. [CrossRef]
22. Li, C. An Efficient Algorithm for Total Variation Regularization with Applications to the Single Pixel Camera and Compressive Sensing. Ph.D. Thesis, Rice University, Houston, TX, USA, 2010. [CrossRef]
23. Yu, W.-K.; Li, M.-F.; Yao, X.-R.; Liu, X.-F.; Wu, L.-A.; Zhai, G.-J. Adaptive compressive ghost imaging based on wavelet trees and sparse representation. *Opt. Express* **2014**, *22*, 7133–7144.
24. Yao, X.-R.; Yu, W.-K.; Liu, X.-F.; Li, L.-Z.; Li, M.-F.; Wu, L.-A.; Zhai, G.-J. Iterative denoising of ghost imaging. *Opt. Express* **2014**, *22*, 24268–24275. [CrossRef] [PubMed]
25. Wang, W.; Wang, Y.P.; Li, J.; Yang, X.; Wu, Y. Iterative ghost imaging. *Opt. Lett.* **2014**, *39*, 5150–5153. [CrossRef]
26. Li, G.; Yang, Z.; Zhao, Y.; Yan, R.; Liu, X.; Liu, B. Normalized iterative denoising ghost imaging based on the adaptive threshold. *Laser Phys. Lett.* **2017**, *14*, 025207. [CrossRef]
27. Wang, W.; Hu, X.; Liu, J.; Zhang, S.; Suo, J.; Situ, G. Gerchberg-Saxton-like ghost imaging. *Opt. Express* **2015**, *23*, 28416–28422. [CrossRef]
28. Zhang, S.; Wang, W.; Yu, R.; Yang, X. Iterative second-order optical eigenmode imaging. *Laser Phys. Lett.* **2016**, *13*, 015003. [CrossRef]
29. Zhou, C.; Wang, G.; Huang, H.; Song, L.; Xue, K. Edge detection based on joint iteration ghost imaging. *Opt. Express* **2019**, *27*, 27295–27307. [CrossRef]
30. Zhang, X.; Meng, X.; Yang, X.; Wang, Y.; Yin, Y.; Li, X.; Peng, X.; He, W.; Dong, G.; Chen, H. Singular value decomposition ghost imaging. *Opt. Express* **2018**, *26*, 12948–12958. [CrossRef]
31. Li, Y.-X.; Yu, W.-K.; Leng, J.; Wang, S.-F. Pseudo-thermal imaging by using sequential-deviations for real-time image reconstruction. *Opt. Express* **2019**, *27*, 35166–35181. [CrossRef]
32. Ruder, S. An overview of gradient descent optimization algorithms. *arXiv* **2017**, arXiv:1609.04747v2. [CrossRef] [PubMed]
33. Huo, Z.; Huang, H. Asynchronous stochastic gradient descent with variance reduction for non-convex optimization. *arXiv* **2016**, arXiv:1604.03584v4.
34. Fliege, J.; Vaz, A.I.F.; Vicente, L.N. Complexity of gradient descent for multiobjective optimization. *Optim. Method. Softw.* **2019**, *34*, 949–959.
35. Du, S.S.; Lee, J.D.; Li, H.; Wang, L.; Zhai, X. Gradient descent finds global minima of deep neural networks. *arXiv* **2019**, arXiv:1811.03804v4. [CrossRef]
36. Liu, H.C.; Yang, H.; Xiong, J.; Zhang, S. Positive and negative ghost imaging. *Phys. Rev. Appl.* **2019**, *12*, 034019.
37. Leng, J.; Yu, W.-K.; Wang, S.-F. Formation mechanism of correspondence imaging with thermal light. *Phys. Rev. A* **2020**, *101*, 033835. [CrossRef]
38. Yu, W.-K.; Leng, J. Probability theory of intensity correlation in ghost imaging with thermal light. *Phys. Lett. A* **2020**, *384*, 126778. [CrossRef]

39. Aslantas, V.; Kurban, R. A comparison of criterion functions for fusion of multi-focus noisy images. *Opt. Commun.* **2009**, *282*, 3231–3242. [CrossRef]
40. Redondo, R.; Bueno, G.; Valdiviezo, J.C.; Nava, R.; Cristóbal, G.; Déniz, O.; García-Rojo, M.; Salido, J.; Fernández, M.M.; Vidal, J.; et al. Autofocus evaluation for brightfield microscopy pathology. *J. Biomed. Opt.* **2012**, *17*, 036008. [CrossRef]
41. Floyd, R.; Steinberg, L. An adaptive algorithm for spatial grey scale. *Proc. Soc. Inf. Display* **1976**, *17*, 75–77. [CrossRef]