

# Color image encryption algorithm based on compressed sensing and DNA

Junfei Yang, Youbao Chang\*, Yang Chen, Song Qian, Long Li  
School of information Engineering, Xinjiang Institute of Technology, Aksu, Xinjiang, China

## ABSTRACT

Based on the combination of compressed sensing principle and DNA, a more effective and secure new method for image augmentation is proposed in this paper. After the original image is compressed by the measurement matrix, the DNA coding rule is used to perform the rule operation with the chaotic sequence generated by the Logistic chaotic system. This step is to perform pixel diffusion on the image, and use the Lorenz chaotic system to construct the index sequence to index the diffused image. Scrambling to obtain an encrypted image, and then decrypting it, using the principle of compressed sensing to process the decrypted image, and through the recovery algorithm (this paper uses the OMP reconstruction algorithm), the original image is successfully restored.

**Keywords:** Image encryption, compressed sensing, DNA encoding

## 1. INTRODUCTION

Since entering the 21st century, the rapid development of global information technology has had a significant impact on industries such as multimedia, computing, and the Internet of Things. This is mainly reflected in people's daily lives, with noticeable changes in work, lifestyle, and learning. People can interact, store, and share multimedia data such as text, images, videos, and voice through cloud-based uploading. In the era of information technology, human productivity has been greatly affected, driving the development of education, economy, healthcare, and scientific research. However, when transmitting and storing private data in the cloud, people often face threats and intrusions from hackers. Therefore, an increasing number of research institutions and scholars are beginning to focus on how to ensure the confidentiality and integrity of private data, which has become an important research direction.

The traditional encryption algorithms, such as the International Data Encryption Algorithm (IDEA), Advanced Encryption Standard (AES), Data Encryption Standard (DES), are mainly used for encrypting text and are not specifically designed for encrypting image information. These traditional algorithms typically transform plaintext data into binary data streams for encryption, without fully considering the inherent properties of digital images such as large data volume, high data redundancy, and strong correlation between adjacent pixels. Therefore, when encrypting digital images, these traditional algorithms have drawbacks such as low efficiency and poor security<sup>1</sup>. As a result, the security and efficiency of image encryption algorithms have become a focal point in the field of information security.

## 2. THEORY OF IMAGE ENCRYPTION

### 2.1 Chaos theory

In the field of chaotic image encryption, chaotic systems have multiple dimensions of chaos, including one-dimensional chaos, two-dimensional chaos, or even higher-dimensional chaotic systems. In this article, we will introduce some commonly seen chaotic systems.

- Logistic mapping

The Logistic mapping has a simple structure, is relatively easy to implement, and possesses excellent characteristics, making it one of the most widely used chaotic systems in the field of image encryption. This chaotic system was proposed by the mathematical ecologist Robert in the 1980s as a one-dimensional chaotic mapping, namely the Logistic map<sup>2</sup>, which is mathematically defined as:

\*1339141501@qq.com

$$x_{n+1} = \mu x_n (1 - x_n) \tag{1}$$

In the equation,  $\mu$  is the system parameters, and  $\mu \in (0, 4)$ ;  $x_n \in (0, 1)$ ,  $n = 1, 2, 3 \dots$

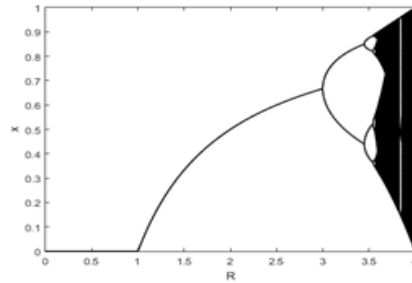


Figure 1. Logistic mapping bifurcation diagram.

Figure 1 is a bifurcation diagram of Logistic mapping, which shows the impact of parameter values on the system state. It can be clearly seen from the figure that when the system is iterated and the parameter value range is between 0 and 1, the system is in a state of no solution, and the value of  $x_{n+1}$  is close to zero at this time. When the system has only one solution, its parameter value is between 1 and 3 and the value of the solution is positively correlated with the value of the parameter. When the number of system solutions becomes two, its parameter value is 3. When the system has two initial stable solutions, its parameter value range is between 3 and 3.569945, and its number will gradually increase as the parameters increase. The periodic bifurcation phenomenon means that when the parameters gradually increase and reach a certain value, the stable solutions of the system will appear one by one, and the number of them will be 4, 8, 16, etc. At the same time, when the system enters a complex chaotic state, its parameter value range is between 3.569945 and 4 and the phenomenon of periodic bifurcation will no longer exist. In addition, when the logistic mapping is in the full mapping state, its parameter value is 4.

- Lorenz Chaos System

The Lorenz chaotic system is a continuous three-dimensional chaotic system discovered in convection experiments. This system is a very typical autonomous dynamical system because the equations of this system are not affected by the system's independent variables, as shown in equations (2).

$$\begin{cases} x' = \alpha y - \alpha x \\ y' = \beta x - xz - y \\ z' = xy - \gamma z \end{cases} \tag{2}$$

In the equation,  $\alpha, \beta, \gamma$ , are system control parameters, when the Lyapunov exponents have positive values and the system is in a chaotic state, their values are 10, 28, and 8/3, respectively. The chaotic attractor is show in Figure 2.

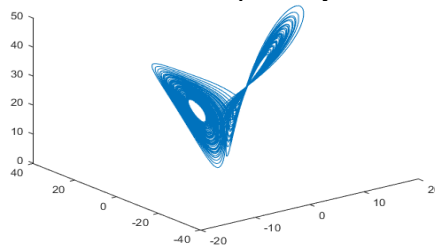


Figure 2. Lorenz chaotic attractor.

- Chen Chaos System

Professor Chen of China conducted further exploration based on the Lorenz chaotic system and subsequently proposed the Chen chaotic system in the late 1990s<sup>3</sup>. Its topological structure is more complex than that of Lorenz. The dynamical equations of the Chen chaotic system are shown in (3).

$$\begin{cases} x' = \alpha y - \alpha x \\ y' = (\gamma - \alpha)x - xz + \gamma y \\ z' = xy - \beta z \end{cases} \quad (3)$$

In the equation,  $\alpha$ ,  $\beta$ ,  $\gamma$  are system control parameters, When the values are 35, 3, and 28, the system will enter a chaotic state at that time. If the initial value is  $x=1, y=1, z=1$ , the chaotic attractor generated is shown in Figure 3.

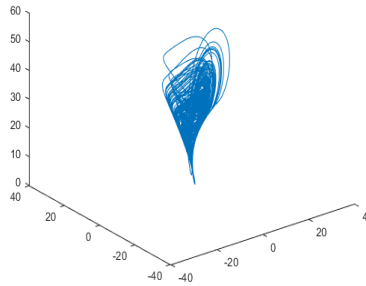


Figure 3. Chen chaotic attractor.

- Chen hyperchaotic system

The Chen hyperchaotic system is based on the Chen chaotic system and is especially extended. Its formula is:

$$\begin{cases} \dot{x} = -\alpha x + \alpha y \\ \dot{y} = -xz + dx + cy - q \\ \dot{z} = xy - bz \\ \dot{q} = x + k \end{cases} \quad (4)$$

Among them,  $a, b, c, d$  and  $k$  are system control parameters. When the system is in a hyper-chaotic state, their values are 35, 3, 28, 16 and  $-0.7 < k < 0.7$ . When  $k=0.2$ , using the above formula, we can calculate the Lyapunov exponents of the hyper-chaotic system. The Lyapunov exponents of this hyper-chaotic system have two positive values, which are 1.552 and 0.023, thus meeting the definition of hyper-chaotic motion. Therefore, this system is in a hyper-chaotic state. Compared to other chaotic systems, the Chen hyper-chaotic system exhibits more complex chaotic behavior because it has more than one positive Lyapunov exponent, making it more suitable for application in image encryption.

In this algorithm design, the Logistic mapping and the Lorenz chaotic system were used to generate chaotic sequences and index sequences, which were then used to perform DNA encryption and image scrambling, ensuring the security of the algorithm.

## 2.2 Basic theory of DNA

DNA computing is a cutting-edge research area that combines biology and computer science, promoting the rapid development of DNA technology. In the 1990s, Dr. Adleman proposed a new research field, DNA computing, at the University of Southern California<sup>4</sup>. Using DNA molecules as a computing medium, he successfully used biological methods to solve the directed Hamiltonian path problem with 7 vertices. The emergence of DNA computing represents a new computing paradigm, breaking the existing mode and opening up a completely new path for solving complex problems. In 1953, Watson and Crick determined through X-ray diffraction experiments that DNA has a double helix structure. Continuous decomposition of the DNA structure can eventually separate nitrogenous bases[5]. DNA contains four different bases: adenine (A), guanine (G), cytosine (C), and thymine (T), often abbreviated as A, G, C, and T. A frequently pairs with T, and C frequently pairs with G. Each component follows the rule that  $A=T$  and  $G=C$ , with an equal number of pyrimidines and purines. The basic principle of DNA computing is derived from mathematics. Each DNA strand is composed of A, T, G, and C, somewhat similar to the “0” and “1” encoding in computers. Based on the inherent double helix structure of DNA and Watson-Crick complementary base pairing rules, it is mapped into a DNA sequence and different biological enzymes are used to act on different operators. Finally, through controllable biochemical reactions, the problem to be solved is formed.

The three operation modes of DNA sequences are addition, subtraction, and exclusive OR (XOR). When any encoding rule is chosen for DNA computation, the result is deterministic. Depending on the chosen encoding rule, the operation

rules of the encryption algorithm will also change, increasing the security of the encryption algorithm. In the process of DNA encryption, pixel values are first encoded into DNA, followed by DNA addition, subtraction, or XOR operations, ultimately resulting in the encrypted image. The decryption process involves reversing the DNA operation rules, such as converting DNA subtraction into DNA addition, and then decrypting it according to the previously chosen encoding rule to restore the original image.

Thymine (T), often abbreviated as A, G, C, and T. A frequently pairs with T, and C frequently pairs with G. Each component follows the rule that A=T and G=C, with an equal number of pyrimidines and purines. The basic principle of DNA computing is derived from mathematics. Each DNA strand is composed of A, T, G, and C, somewhat similar to the “0” and “1” encoding in computers. Based on the inherent double helix structure of DNA and Watson-Crick complementary base pairing rules, it is mapped into a DNA sequence and different biological enzymes are used to act on different operators. Finally, through controllable biochemical reactions, the problem to be solved is formed.

The three operation modes of DNA sequences are addition, subtraction, and exclusive OR (XOR). When any encoding rule is chosen for DNA computation, the result is deterministic. Depending on the chosen encoding rule, the operation rules of the encryption algorithm will also change, increasing the security of the encryption algorithm. In the process of DNA encryption, pixel values are first encoded into DNA, followed by DNA addition, subtraction, or XOR operations, ultimately resulting in the encrypted image. The decryption process involves reversing the DNA operation rules, such as converting DNA subtraction into DNA addition, and then decrypting it according to the previously chosen encoding rule to restore the original image.

### 2.3 Compressed sensing theory

Compressed sensing consists of three parts: sparse representation, measurement matrix, and reconstruction algorithm.

- Sparse representation

Only when the signal has the characteristics of sparsity ( $K \ll M \ll N$ ), it is possible to recover the original signal of length  $N$  from the observed  $M$  measurement values through  $K$  larger coefficients. If a one-dimensional signal  $x \in \mathbb{R}^N$ , sparse basis  $\Psi = [\psi_1, \psi_2, \dots, \psi_N]^T$  ( $\Psi \in \mathbb{R}^{N \times N}$ ) sparse representation of signals  $x$ . Then the transformation coefficient  $\alpha$  can be expressed as:

$$\alpha = \Psi^T x \tag{5}$$

In equation (5),  $\Psi$  represents the orthogonal basis, while  $x$  and  $\alpha$  respectively represent the representation of the same signal in different domains (such as time domain and frequency domain). The main characteristic of this transformation method is that most of the transform coefficients are close to zero, and only a very small number of transform coefficients contain the main information of the signal. In compressive sensing, the prerequisite for whether a signal can be sensed compressively is whether the signal is sparse. However, truly sparse signals are rare in the real world, which limits the application of compressive sensing. Fortunately, it has been found that when most signals in the time domain are projected onto a certain transform domain, the resulting high-bit signals can be similar to sparse, that is, compressible signals. Therefore, if we want compressive sensing to effectively compress and sample data, we only need to find the optimal sparse basis corresponding to the characteristics of the data itself to enable compressive sensing to effectively compress and sample it.

- Measurement matrix

The term “Measurement Matrix” is usually translated as “Measurement matrix” in Chinese, occasionally it can also be translated as “Observation matrix” or “Sampling matrix”. In compressive sensing, the form of the measurement matrix is very important because the design of the measurement matrix is the main means of obtaining the signal, and the performance of the measurement matrix can directly affect the reconstruction algorithm of the signal. For signals  $x$ , the measurement matrix  $\Phi$  ( $\Phi \in \mathbb{R}^{M \times N}$ ,  $M \ll N$ ) can be used for linear projection to obtain linear measurement values.

$$y = \Phi x = \Phi \Psi \alpha = \Theta \alpha \tag{6}$$

In equation (6), the measurement value  $y \in \mathbb{R}^{M \times N}$ ,  $M \ll N$  is used to reduce the measurement value of the measurement object from  $N$  dimensions to  $M$  dimensions, thereby completing the compression process while sampling the original signal. The requirement for designing the measurement matrix is that the  $K$  measurement values also contain all signals of the original signal, and the original signal will not be destroyed, thus ensuring the accurate reconstruction of the original signal. The original signal can be reconstructed using the sensing matrix  $\Theta = \Phi \Psi$ . However, since there are more unknowns in the system

of equations than there are equations, a definite solution cannot be obtained, resulting in the signal being unable to be reconstructed. If the original signal  $x$  is  $K$ -sparse and  $\theta$  follows the finite isometric property (RIP), then only  $M$  measurements using  $K$  coefficients are needed to accurately reconstruct the original signal.

The mathematical definition of RIP is:

$$(1-\delta_K)\|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1+\delta_K)\|x\|_2^2 \quad (7)$$

- Reconstruction algorithm

In compressive sensing theory, the role of the signal reconstruction algorithm is to recover the compressed observed values into sparse signals. Therefore, the reconstruction algorithm is an important component of the theory, and its quality directly affects the practical application of compressive sensing theory.

The algorithm for reconstructing the signal also includes the exploration of the norm  $l_0$  and  $l_1$ .

$$\min \|\alpha\|_0, \text{ s.t. } \Phi \alpha = y \quad (8)$$

If this problem is solved by exhaustive attack, it is an NP-hard problem. Typically, we use convex optimization methods to try to solve it, i.e

$$\min \|\alpha\|_1, \text{ s.t. } \Phi \alpha = y \quad (9)$$

The greedy algorithm, convex optimization algorithm, and composite algorithm can be used to reconstruct the signal into the original signal, and the three commonly are used reconstruction algorithms. Among them, the greedy algorithm includes matching pursuit algorithm, orthogonal matching pursuit (OMP) algorithm, and segmental orthogonal matching pursuit algorithm. Convex optimization algorithms include threshold iteration and basis pursuit. Composite algorithms mainly include Fourier sampling and chain pursuit method. These algorithms can be flexibly selected according to specific situations for signal reconstruction and restoration of original information.

### 3. ALGORITHM IMPLEMENTATION

#### 3.1 Encryption process

- Image compression

First, the  $256 \times 256$  color lena image that needs to be encrypted is separated by RGB, and the R, G, and B components obtained after separation are used in turn to use the measurement matrix generated by the Gaussian random matrix. The size of the measurement matrix  $\Phi$  is  $230 \times 256$  (here We need to explain that the random number seed is inserted and the specified measurement matrix is generated with the random number seed. We need to remember this random number seed. This seed can be used to generate the same observation matrix during decryption) and compress it into the original 0.9 times the size. At this time, the data we get after compression is a floating point number and contains negative numbers. In order to facilitate later encryption and decryption, we convert it into an integer in the range of 0-255. The compressed image is shown in Figure 4.

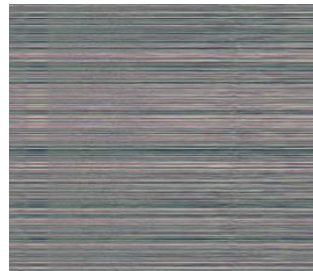


Figure 4. Compressed image.

- DNA coding

The compressed image size is  $230 \times 256$ , and the total number of data is  $230 \times 256 = 58880$ . Next, the data obtained after image compression is stretched into a column to obtain  $58880 \times 1$  data. Each data is converted into an 8-bit binary bit for representation, and  $58880 \times 8$  data is obtained. A DNA encoding operation is performed for every 2 bits of data, which

means that each pixel value requires 4 encoding operations<sup>6</sup>. We convert the new data obtained after encoding into a binary matrix with a size of 58880\*8. The image after DNA encoding is shown in Figure 5.



Figure 5. Image after DNA encoding.

- DNA rule operation (diffusion)

Use Logistic chaos to generate a chaotic sequence, generate a sequence  $x_s$  of length 58880, and convert the generated chaotic sequence into an integer ranging from 0 to 255<sup>7</sup>. Next, we use the converted chaotic sequence to perform DNA encoding operations, and also convert it into a binary matrix and the binary matrix converted after DNA encoding to perform DNA rule operations. Here we use DNA addition operation, that is, each group of 2 bits is converted into a chaotic sequence. If the binary matrix is added, we will get the DNA-encrypted data. This step is diffusion. The data encrypted by DNA is a 58880\*8 binary 01 matrix. By converting it from binary to decimal and merging it, we can get our DNA-encrypted image with a size of  $M$ \*width. The image of DNA rule operation is shown in Figure 6.



Figure 6. Image after DNA operation.

- Scrambling

The Lorenz chaos system is used to scramble the DNA-encrypted image. We specify the initial value of the Lorenz chaos system and remove the first 1,000 items to obtain better randomness. We sort the generated chaotic sequences according to the original values. The positions in the sequence are constructed with index scrambling, and the pixels of the DNA-encrypted image are scrambled, and finally our final encrypted image is obtained. The scrambled image is shown in Figure 7.

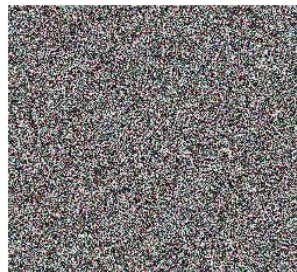


Figure 7. Final encrypted image.

### 3.2 Decryption process

The image decryption process can be seen as the opposite process of image encryption. In the whole process of decryption, the encoding and decoding methods of DNA correspond to the methods of DNA decoding and encoding in the encryption process. The encoding rules must be selected consistent with the encryption process, and the addition operation of DNA

must be replaced with the corresponding ones. The corresponding subtraction operation. For the chaotic sequence generated by the chaotic system, we only need to control its parameters and the method of initial value processing to remain unchanged. What needs to be noted here is the generation of the measurement matrix. Because a random number seed is used to generate the specified measurement matrix during the encryption process, the same random number seed needs to be used here to generate the same measurement matrix, and then the original image Perform DCT sparse processing, and linearly project the data obtained after sparse processing onto the Gaussian random measurement matrix, then the linear measurement values can be obtained. Finally, the OMP algorithm is used to accurately reconstruct the original image information. The final decrypted image is obtained by reconstructing the images of each channel and then merging them. The decrypted image is shown in Figure 8.



Figure 8. Decrypted image.

### 3.3 Image encryption result analysis

Figure 9 presents the grayscale histogram of the RGB component of the original image and the corresponding grayscale histogram of the encrypted image.

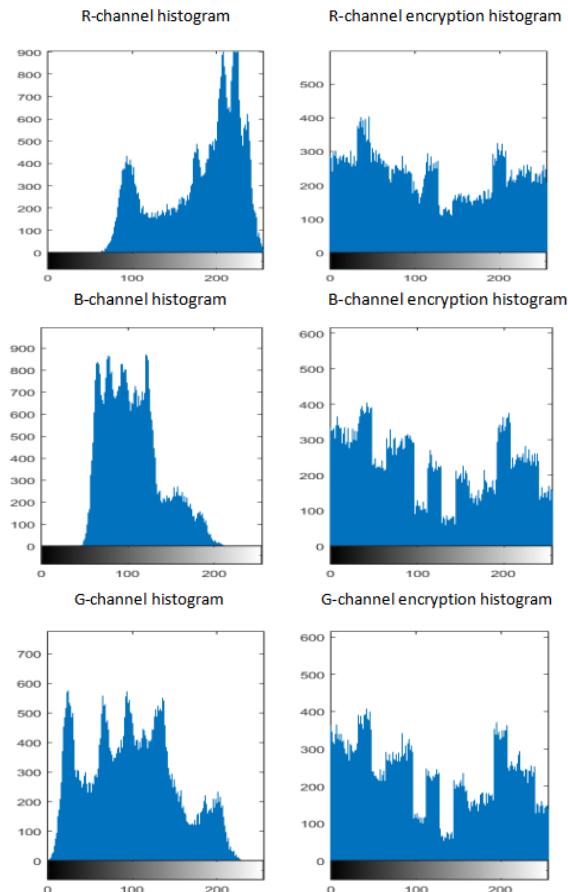


Figure 9. Grayscale histogram of RGB components of the original image.

It can be intuitively found from Figure 9 that the grayscale histogram distributions of the RGB components of the original image are all different, but the grayscale histograms of the RGB components of the encrypted image are very similar. Therefore, it can be judged and analyzed that the image encryption algorithm in this article can effectively withstand statistical analysis attacks.

The gray value in the image can be measured by information entropy to determine its random distribution, which is an important condition for judging randomness. When the information entropy of an image is greater, the distribution of gray values is more uniform and closer to the ideal value, and the less effective information criminals can obtain. When the information entropy of an image is lower, it is more susceptible to tampering and malicious attacks by criminals. Next, we will analyze the information entropy of the encrypted image obtained by running the algorithm in this article to determine the security of the encryption algorithm.

R-channel original information entropy: 7.2432	Encrypted image information entropy: 7.9458
B-channel original information entropy: 7.5774	Encrypted image information entropy: 7.8908
G-channel original information entropy: 6.9223	Encrypted image information entropy: 7.8890

Figure 10. Comparison of information entropy.

From the analysis of the data results obtained in Figure10, it can be seen that the image information entropy value obtained after encryption by the image encryption algorithm proposed in this article is close to 8, which is in line with the ideal situation. This means that the degree of chaos in the encrypted image is as close as possible to the ideal value of 8, which proves that this algorithm can generate encrypted images with more random pixel distribution, higher security and strong resistance to attacks.

#### 4. CONCLUSIONS

At the beginning of this experiment, RGB separation was performed on the original image, and the three separated RGB components were compressed through the generated measurement matrix, and then DNA encoding operations were performed on them. The chaotic sequence generated by the Logistic chaos system was also subjected to DNA encoding operations. Then the two binary matrices after the DNA encoding operation are subjected to DNA rule operations. This step is the diffusion operation. Next, the Lorenz chaos system is used to perform an index scrambling operation on them to obtain the scrambled encrypted image, and then decrypt it. operation, and finally complete the restoration of the original image through the relevant knowledge of compressed sensing principles.

Finally, we will verify various indicators of the encrypted image, including correlation, key space, histogram, information entropy, etc., obtain the experimental results and analyze them. Through the experimental data obtained, we proved that this article The security and reliability of the algorithm, as well as the large key space. Its performance meets the current demand for information security and can be used in the field of image encryption to provide sufficient protection for our information security.

#### REFERENCES

- [1] Jie, F., Ping, P. and Zeyu, G., "A meaningful visually secure image encryption scheme," 2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService), (2019).
- [2] May, R. M., "Simple mathematical models with very complicated dynamics," *Nature*, 261(5560), (1976).
- [3] Chen, G., et al., "Yet another chaotic attractor," *International Journal of Bifurcation and Chaos*, 9(7), (1999).
- [4] Adleman Leonard, M., "Molecular computation of solutions to combinatorial problems," *Science*, 266(5187) (1994).
- [5] Liu, C., Yang, J., et al., "Chaotic message encryption algorithm based on DNA strand substitution reaction," *Journal of Changchun Normal University*, 42(02), 65-71(2023).
- [6] Tang, C., [Design and Analysis of Digital Image Encryption Algorithm Based on Improved Chaotic System], Fuyang Normal University, (2023).
- [7] Xu, A., [Unified Image Encryption Algorithm Based on Chaotic System and Wavelet Transform], Jiangxi University of Finance and Economics, (2023).