applied sciences

MDPI

Article

A Novel Image Encryption Algorithm Based on Voice Key and Chaotic Map

Jing Li [1,*], Tianshu Fu [2], Changfeng Fu [3,*] and Lianfu Han [3]

[1] College of Humanities and Social Sciences, Heilongjiang Bayi Agricultural University, Daqing 163319, China
[2] School of Physics and Electronic Engineering, Northeast Petroleum University, Daqing 163318, China; 090206@nepu.edu.cn
[3] College of Electronical and Information Engineering, Changshu Institute of Technology, Changshu 215506, China; lianfuhan@nepu.edu.cn
* Correspondence: lijing1953125@byau.edu.cn (J.L.); changfengfu@nepu.edu.cn (C.F.); Tel.: +86-186-452-99670 (J.L.); +86-158-458-28703 (C.F.)

Featured Application: In this article, an encryption algorithm is proposed that can be used for remote identity authentication. The specific application method is described in the third section of this article. Initially, the user's voice information is extracted as data, which is subsequently used as the key for image encryption. The encrypted image can be successfully decrypted during identity verification only when the user's voice data are correct, which leads to successful identity verification. This is a new application scheme, and no similar application method has been described in the existing literature, to the best of our knowledge.

Abstract: This paper proposes a new image encryption algorithm. First, time-domain and frequency-domain features of the user's voice are extracted to generate a voice key. Second, the key is iterated through a chaotic map multiple times to map the key data to the chaotic oscillation region, and, subsequently, the parameters of the oscillation area are used to encrypt the user's image. Third, at the time of decryption, the user's latest voice data are re-extracted to generate a new voice key and decrypt the encrypted image. The encrypted image cannot be successfully decrypted if there are differences between the two extracted voices in the time or frequency domain. Finally, the experiments are performed using 80 groups of face images and voice data, all of which pass the encryption and decryption experiments. In addition, various safety tests have been carried out on the algorithm. The key sensitivity of the algorithm is verified by the normalized cross-correlation parameter C_{ncc}. The effective anti-attack ability of the algorithm is verified by measuring the correlation between adjacent pixels, the number of changing pixel rate (NPCR) and the unified averaged changed intensity (UACI). The key space of the proposed algorithm is greater than 2^{100}, and it has good anti-cracking ability.

Keywords: chaotic map; image encryption; voice key; authentication

check for updates

Citation: Li, J.; Fu, T.; Fu, C.; Han, L. A Novel Image Encryption Algorithm Based on Voice Key and Chaotic Map. *Appl. Sci.* **2022**, *12*, 5452. https://doi.org/10.3390/app12115452

Academic Editor: David Megías

Received: 16 April 2022
Accepted: 24 May 2022
Published: 27 May 2022

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

Copyright: © 2022 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

The purpose of this work is to propose a novel image encryption algorithm that can be used to improve the reliability of remote identification. There are four major contributions of this work. First, a method for generating keys based on speech processing technology is proposed. Second, a method for image encryption based on chaotic maps is proposed. Third, a new set of image encryption equations are constructed. Finally, a scheme for applying image encryption technology to remote identity authentication is proposed. This work aims to present an identity authentication scheme for the elderly to receive social security funds while residing in different places. In our region (China), the elderly are required to confirm their identity on the internet before receiving pensions. The scheme proposed in this work can prevent someone from using virtual dynamic portrait technology to defraud the social security funds. The above work is realized based on experiments and compared with the experimental results presented in other works.

1.1. Analysis of Common Identity Authentication Methods

Broadly, biometric systems can be divided into two main types: unimodal and multimodal biometric systems. Unimodal systems are based on using a single source of information to establish the person's identity, for example, iris, retina, face, speech, fingerprint, finger-vein, signature, etc. The former six methods are based on biometrics, i.e., the use of various biological characteristics of the human body to achieve identity authentication. The last method, i.e., signature recognition, is a kind of behavior authentication method [1].

Many methods have been presented for identification and verification based on any of the above information sources, where every method employs a different strategy. A review of some prominent solutions is now presented.

Lim et al. in [2] proposed an efficient method for personal identification by analyzing iris patterns, which have a high level of stability and distinctiveness. To improve the efficiency and accuracy of the proposed system, they presented a new approach to making a feature vector compact and efficient by using wavelet transform, and two straightforward but efficient mechanisms for a competitive learning method, utilizing a weight vector initialization and winner selection.

A method for human identification based on retinal images is presented in [3]. The proposed system is composed of two main parts: a feature-extraction component and a decision-making component. In the feature-extraction component, blood vessels are first extracted and then thinned by a morphological algorithm. Then, two feature vectors are constructed for each image, by utilizing angular and radial partitioning. In the article, a fuzzy system with Manhattan distances of two feature vectors as the input and a similarity measure as the output was added to the decision-making component.

An exhaustive multifactor face authentication system using a neuro-fuzzy approach is described in [4]. During the enrollment process, facial region of a still image of the authorized user is captured and features are extracted using a local tetra pattern (LTrP) technique. The features are given as the input to the neural network, namely a fuzzy adaptive learning control network (FALCON), for training and classification of features. During the authentication process, an image that can vary with expression, pose, illumination and occlusion factors is taken as a test image and the test image is inputted to the LTrP and FALCON to train the system based on the features of the test image. Then, these trained features are compared with an existing feature set using the newly proposed multifactor face authentication algorithm to authenticate a person.

Hyun Park and Tae Guen Kim, in [5], develop a voice-based authentication model that learns and discriminates each user's voice data using a deep neural network. In addition, they also present a synthesis speech detection method that is used to prevent a masquerading attack using synthetic voices. The proposed method can be divided into two modules: an MFCC based user authentication module and a Mel-Spectrogram-based synthetic speech detection module.

A method for signature verification and recognition is presented in [6], which represents signature verification and recognition using zone wise statistical features. During the first phase, a knowledge base is constructed by training samples using the zone wise statistical features. During the second stage, i.e., the testing phase, the processed image is obtained with zoning wise statistical features and the signature is recognized using neural network classifiers. MATLAB was used to design this signature recognition and verification system.

Boucherit et al., in [7], present a new approach based on a deep learning model to achieve personal identification through finger vein patterns. They employed an improved deep network, named Merge Convolutional Neural Network (Merge CNN), which uses several CNNs with short paths. The scheme is based on the use of multiple identical CNNs with different input image qualities, and unification of their outputs into a single layer.

Brain wave recognition is a new biometric authentication method. The scheme performs authentication and identification by collecting EEG signals during behavioral and/or mental activity. This scheme presents state-of-the art solutions and recommendations for addressing security and privacy problems by proposing a novel, EEG-driven, secure and reliable cognitive authentication system for an IoT-based healthcare system. Please note that the brain wave authentication has wide application prospects, high reliability and uniqueness. However, promotion of this technology is slow due to the high cost of brain–computer interface devices and the complexity of the software algorithms [8].

A dynamic application-partitioning workload task-scheduling-secure (DAPWTS) algorithm was proposed in 2021. It consists of different schemes, such as a min-cut algorithm, a searching node, energy-enabled scheduling, failure scheduling and security schemes. The goal of this methods is to minimize the energy consumption of the nodes and divide the application between local nodes and edge nodes by applying the secure min-cut algorithm [9]. If the scheme proposed in this work were generalized, we could learn from its technical framework to optimize the design of identity authentication methods.

Although these systems have been widely employed in government and civilian sensitive applications with a high level of security, they often suffer from a number of critical limitations and problems that can affect their reliability and performance. With the exceptions of the fingerprint identification method and the brain waves method [10], these biometric methods do not require any direct physical contact by the user, thus making them convenient for the user. The iris recognition, retina recognition and finger-vein recognition methods have high reliability and robustness against forgery [11]. However, they also suffer from disadvantages, such as a high equipment cost, large volume, noise, intra-class variations and inter-class similarities [12,13]. Moreover, retina recognition can even affect the health of the user [14,15].

All these drawbacks of unimodal systems can be efficiently addressed by systems combining evidence from multiple sources of information to identify a person's identity; these are referred to as multimodal systems. Multimodal systems can produce sufficient population coverage by efficiently addressing problems related to the enrollment phase, such as non-universality. Furthermore, these systems can provide higher accuracy and a greater resistance to unauthorized access by an imposter than unimodal systems, due to the difficulty of spoofing or forging multiple biometric traits of a legitimate user at the same time. The most fundamental issue for the designer of the multimodal system is choosing the most powerful biometric traits from multiple sources to be incorporated within the system, and finding an efficient method of fusing them [16]. The following are some representative multimodal system authentication schemes.

Wu et al. in [17] proposed an identity authentication framework, which can extract and verify personal information through face verification and ID image recognition. The identity authentication is realized using the proposed face verification model, which is called Inception-ResNet Face Embedding (IRFE). IRFE uses an Inception-ResNet structure to ensure a good feature extraction, aiming at accurate face verification. Moreover, a robust ID card extraction method named Morphology Transformed Feature Mapping (MTFM) is proposed to extract ID information.

Navya Saxena and Devina Varshney, in [18], propose a holistic solution for the implementation of Smart Home Security, which helps in improving privacy and security by using two independent and emerging technologies of facial authentication and speech recognition. This method involves facial recognition by taking a real-time feed of the person at the door; analysis of the live feed is then conducted, in which the face recognized is authenticated by comparing it with data regarding owners in a database, thereby matching the face to a name. Speech recognition was used to double check the output of facial authentication. This entire process is carried out with the help of neural networks. If there is an unauthorized person at the door, an alert is triggered, and the owner receives a notification of this unauthorized access; they can then choose whether they want to add the person to their database or not. Face recognition is widely used in identity recognition as it is a non-contact approach that uses compact equipment. It is the fastest growing technology among all the relevant technologies. However, a few hidden loopholes have emerged gradually. An audio-driven facial video synthesis, such as Motionface, can make an arbitrarily selected face image correspond to a dynamic image of speech as long as you input a piece of speech. By synthesizing face video from audio and facial images, these deep fake technologies can realize a face swap function, i.e., one face in an image or video can be replaced with another face. These technologies pose a significant threat to the network and personal information security, as well as to the reliability of identity verification based on facial recognition technology. Therefore, we chose to combine facial recognition with speech recognition through the image encryption technology proposed in this article to improve the reliability of identity recognition. The method proposed in this work is used for personal identity authentication, without using any special authentication equipment.

1.2. Combination of Image Encryption and Identity Recognition Technologies

The image encryption utilized in the present study mainly includes encryption technology based on frequency and spatial domains. The encryption technology based on the frequency domain is mainly used for image compression, such as JPEG compression and compression technology based on the wavelet transform. The encryption technology based on the spatial domain is often used in security applications. The use of chaotic maps is common for spatial image encryption. Various existing image encryption algorithms can encrypt face images to ensure the security and privacy of the images, but they cannot be used to improve the reliability of facial image authentication. In order to deal with this issue, an image encryption algorithm based on chaotic maps and an algorithm for extracting human voice features into data are proposed in this article. The time-domain and frequency-domain features from a human voice are extracted to generate keys, which are used for image encryption. For identity recognition, the user's specific voice information is used to decrypt the encrypted image, and then the face recognition can be performed. This encryption and decryption method increases the difficulty of cracking the authentication scheme. The solution can also be applied in various other face recognition authentication situations to improve authentication security.

2. Principles of the Proposed Method

The image encryption principle used here involves two aspects, including key extraction and encryption and calculation of the image. The key extraction adopts two methods involving time-domain and frequency-domain processing of voice information. The key information must not only reflect the user's voice characteristics, but also conform to the oscillation conditions of the chaotic maps so that the encryption result is sensitive enough to the key information.

2.1. Voice Key Generation Process

The user's voice-feature information should be extracted as the key for image encryption prior to encryption. The speech signal is a non-stationary time-varying signal, whose characteristics can be analyzed and described in both the time and frequency domains. Here, sets of time-domain and frequency-domain feature parameters are extracted to perform the image encryption operation. The time-domain analysis method is the simplest and most intuitive analysis method. The time-domain speech parameters include short-term average energy, short-term average zero-crossing rate and short-term autocorrelation function. We take the number of peaks of a single-syllable single waveform as the extracted time-domain parameter. It has a corresponding relationship with the average zero-crossing rate and reflects the difference between various dissimilar syllables. The relevant expression is given in Equation (1).

$$\begin{cases} n_w = \frac{\sum_{m=L_a}^{L_a+L_b} |sgn[z(m)] - sgn[z(m-1)]|}{M \times 4} \\ sgn[z(m)] = \begin{cases} 1 & z(m) \geq 0 \\ -1 & z(m) < 0 \end{cases} \end{cases} \quad (1)$$

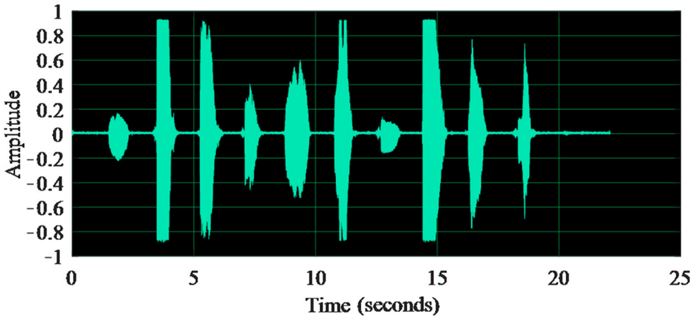
In Equation (1), L_a and L_b are the starting and ending positions of a single syllable in the time domain, respectively; $z(m)$ is the amplitude of the audio position m; and M is the number of single waveforms of a single syllable. A total of six n_w values calculated by Equation (1) are represented by n_w1 to n_w6, and these six values participate in the subsequent chaotic encryption calculation.

The frequency-domain description of voice information is an important feature that reflects the distribution of voice information over different frequencies. Different people have different spectrum distributions of the same syllable. This can be used to distinguish the difference between the pronunciations of various people. Here, discrete fourier transform (DFT) is used to extract the user's voice frequency-domain features. It is a simple and practical method to analyze the voice frequency spectrum characteristics. Its formula is given by Equation (2), as follows:

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j\frac{2\pi}{N}kn} (k = 0,\ 1,\ 2 \ldots N-1) \quad (2)$$

In Equation (2), $X(k)$ represents the data after DFT transformation; $x(n)$ is the nth signal sample; N is the total number of Fourier transform points; and k is the ordinal number of the abscissa of the spectrum, which represents the kth spectrum position. One of the peak frequency values can be selected to participate in the subsequent chaotic encryption calculation. This selection can be based on either a sequential or a magnitude relationship.

As illustrated in Figure 1, in the first phase of the speech features extraction, the characteristic single waveform corresponding to each single-syllable speech datum is extracted. Subsequently, the single waveform is used to synthesize the repetitive long-term signal waveform, such as a waveform with a synthetic duration of three seconds. The waveform undergoes DFT transformation to extract its frequency-domain data, where the frequency data corresponding to the peak amplitude are considered. In addition, the time-domain feature data of the single-syllable characteristic single waveform are extracted, where the number of crests of the single waveform and the total number of single syllables are considered.



(**a**)

Figure 1. *Cont.*

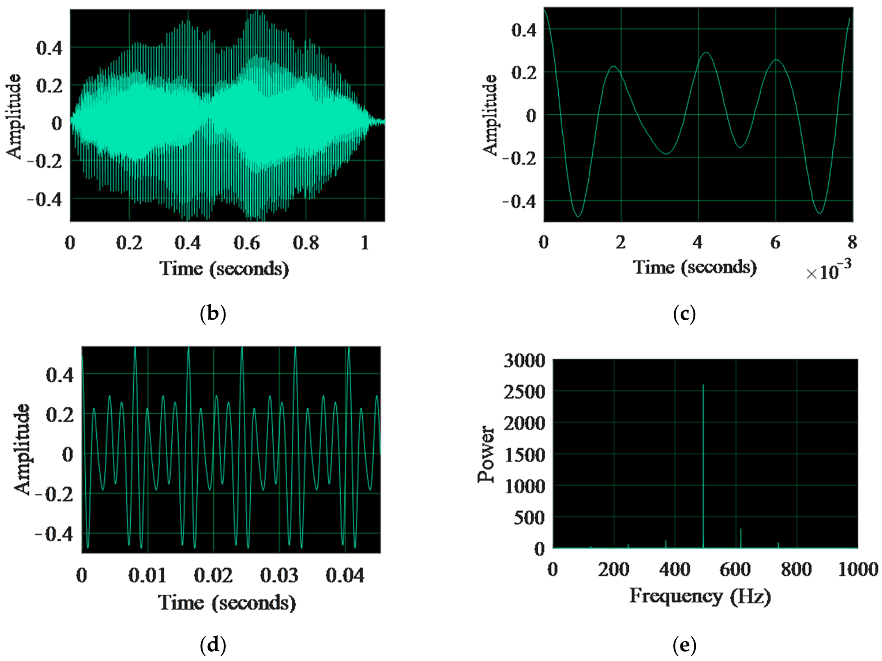


Figure 1. Speech feature extraction. (**a**) Original voice. (**b**) Step1: Single-syllable audio; (**c**) Step2: Single waveform of single-syllable audio; (**d**) Step3: Repetitive synthetic audio of a single waveform; (**e**) Step4: DFT conversion of a single waveform repeatedly synthesized for three seconds of audio.

2.2. Proposed Image Encryption and Decryption Algorithm

2.2.1. Encryption Process

As shown in Figure 2, the image can be encrypted after obtaining the voice key information used for encryption. Here we use the chaotic map for encryption. Chaotic mapping has a wide range of applications in image encryption. The output of a chaotic system that has iterated a certain number of times is sensitive to the initial value and exhibits randomness. Image encryption can take advantage of this sensitivity and randomness. In this paper, we use logistic chaotic mapping to iterate the chaotic system based on the voice key as shown by Equation (3), which is used to calculate the two initial data sets required by the image displacement formula [19].

$$\begin{cases} L_{nL1+1} = L_{nL1} \times \mu_1 \times (1 - L_{nL1}) \\ L_{nL1} = n_w 1/k1 \\ \mu_1 = k2 + n_w 3/k3 \\ L_{nL2+1} = L_{nL2} \times \mu_2 \times (1 - L_{nL2}) \\ L_{nL2} = n_w 2/k4 \\ \mu_2 = k5 + n_w 4/k6 \end{cases} \tag{3}$$

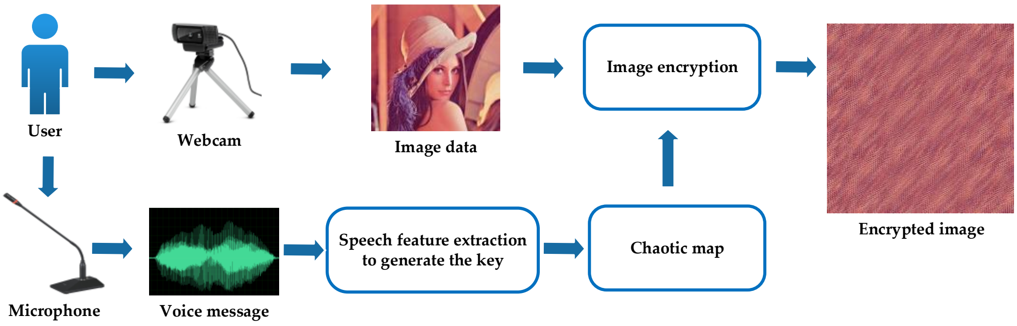


Figure 2. Encryption process.

In Equation (3), L_{nL} is the generated chaotic data; μ_1 and μ_2 are the logistic map parameters; and $k1$ to $k6$ are the preset parameters. According to the chaotic mapping conditions, the initial values of L_{nL1} and L_{nL2} should be between zero and one, and μ_1 and μ_2 can be between 3.5699456 and 4. In addition, the number of iterations for the chaotic map must be sufficient to allow the output data to enter the random oscillation region as shown in Figure 3.

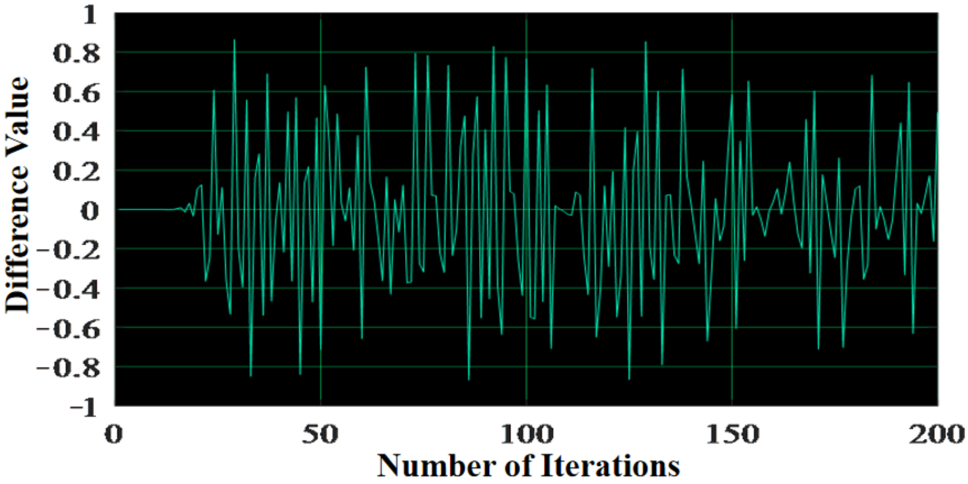


Figure 3. The output of logistic map.

In Figure 3, μ is set to 3.9; the initial values of the two oscillations are selected as 0.100001 and 0.100002; the ordinate represents the difference between the two chaotic data sets; and the abscissa is the number of iterations of the chaotic map. It can be observed from the output that when the number of iterations is greater than 20, the chaotic map enters a random oscillation zone. When performing image encryption, we set the values of $k1$ to $k6$ and the number of iterations after determining the voice key to ensure that the output data are in the random oscillation area. Next, it is necessary to map the iteratively calculated value in Equation (3) to the value range corresponding to the encryption formula. The mapping relationship is shown in Equation (4).

$$\begin{cases} m_1 = round((L_{nL1}(n_w5 + k9 \times f_m + k7) + 1)/2 \times s) \\ b = round((L_{nL2}(n_w6 + k9 \times f_m + k8) + 1)/2 \times s) \end{cases} \tag{4}$$

In Equation (4), L_{nL1} and L_{nL2} are the vectors generated by the iterative Equation (3); round(.) is the rounding function; f_m is the peak frequency obtained after the DFT transformation of the synthesized speech waveform; s is the number of lines of the image to be encrypted; and $k7$ and $k8$ are two adjustable parameters. A slight change in the values of $k7$ and $k8$ will cause a significant change in the encryption result. After each identity authentication, the $k7$ and $k8$ can be reset to strengthen the system security. The parameter $k9$ is an integer multiple of either one or ten. The capacity of the key space can be increased by increasing the resolution of f_m and adjusting $k9$. Algorithm 1 describes the process of image encryption.

$$\begin{cases} f(i_1) = f(j_1)\ i_1 = 1,\ 2,\ \ldots m_1\ j_1 = n_1 - m + 1,\ \ldots, n_1 - 1,\ n_1 \\ f(i_1) = f(j_1)\ i_1 = m_1 + 1,\ \ldots,\ n_1 - 1,\ n_1\ j_1 = 1,\ 2,\ \ldots,\ n_1 - m_1 \\ m_{1m+1} = m_{1m} + b \end{cases} \tag{5}$$

Algorithm 1: Image encryption

```
Input: ($n_w1$~$n_w6$, $k1$~$k9$, $f_m$, image matrix $a$)
begin
    for ($i$ = 1, 3000) do
        $L_{nL1} = n_w1/k1$
        $\mu_1 = k2 + n_w3/k3$
        $L_{nL1+1} = L_{nL1} \times \mu_1 \times (1 - L_{nL1})$
        $L_{nL2} = n_{w2}/k4$
        $\mu_1 = k5 + n_w4/k6$
        $L_{nL1+1} = L_{nL1} \times \mu_1 \times (1 - L_{nL1})$
    end
    $m_1 = round((L_{nL1}(n_w5 + k9 \times f_m + k7) + 1/2 \times$ the number of rows)
    $b = round((L_{nL2}(n_w6 + k9 \times f_m + k8) + 1/2 \times$ the number of columns)
    for ($i$ = 1, the number of rows) do
        for ($j$ = 1, the number of columns) do
            matrix scrambling
        end
    end
    return $a$
end
```

Equation (5) is the formula for vector right shift and down shift scrambling of the image data: m_1 is the number of shifted bits and b is the step length that the next row or column shift number should accumulate after each shift. When the image is encrypted, L_{nL1} and L_{nL2} perform multiple iterative calculations, and the calculation results are subsequently mapped to obtain the two parameters m_1 and b required by the scrambling formula. In the calculation, the numbers of single-waveform peaks n_w1 and n_w2 corresponding to the first and second single syllables of the speech signal, respectively, are correlated with the initial value of the chaotic map. The parameters n_w3 and n_w4 are correlated with the logistic mapping parameter μ, and the peak frequencies of n_w5, n_w6 and audio DFT transformation are used to map the value calculated by chaotic map to the range of m_1 and b. In this way, the image encryption can be completed. Algorithm 2 describes the process of image matrix scrambling.

Algorithm 2: Matrix scrambling

```
Input: (matrix $a$, $m_1$, $b$)
begin
    for ($i$ = 1, the number of rows) do
        $m_1$,=mod($m_1$, the number of rows)
        Left cyclic shift ($m_1$)
        $m_1 = m_1 + b$
    end
    for($j$ = 1, the number of columns) do
        $m_1$,=mod($m_1$, the number of columns)
        Downward cyclic shift ($m_1$)
        $m_1 = m_1 + b$
    end
    return $a$
end
```

2.2.2. Decryption Process

As shown in Figure 4, the image decryption is the reverse process of image encryption. The chaotic mapping formula in the decryption process is exactly the same as that in the encryption process. The shift formula for image decryption is shown in Equation (6). Furthermore, the code execution sequence for shifting image rows and columns in a decryption program is the opposite of that in the image encryption.

$$\begin{cases} f(i_1) = f(j_1)\ i_1 = n_1 - m + 1, \ldots, n_1 - 1, n_1\ j_1 = 1,\ 2,\ \ldots m_1 \\ f(i_1) = f(j_1)\ i_1 = 1,\ 2,\ \ldots,\ n_1 - m_1\ j_1 = m_1 + 1,\ \ldots,\ n_1 - 1,\ n_1 \\ m_{1m+1} = m_{1m} - b \end{cases} \quad (6)$$

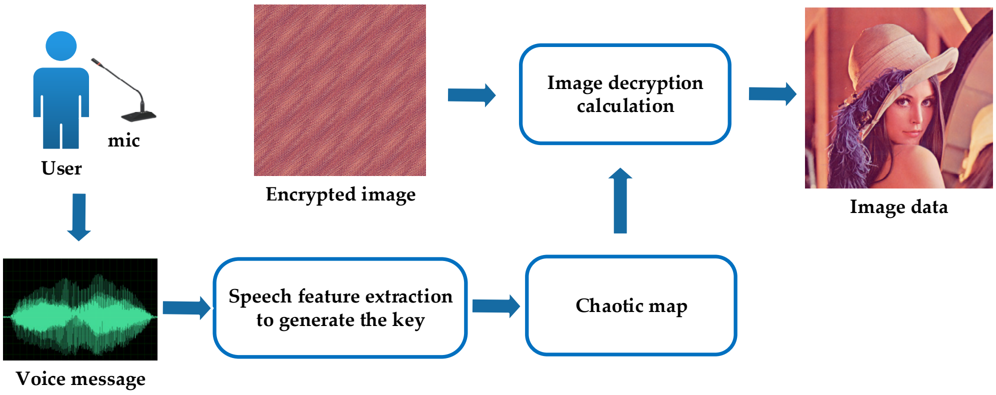

User
mic
Encrypted image
Image decryption calculation
Image data
Voice message
Speech feature extraction to generate the key
Chaotic map

Figure 4. Decryption process.

3. Application of the Proposed Image Encryption Algorithm in Identification Technology

A novel application scheme is proposed in this paper. To the best of our knowledge, no similar application scheme has been proposed in the existing literature. The process of identity recognition can be designed on the basis of the encryption and decryption algorithms described in Section 2. As shown in Figure 5, it consists of several steps: (1) In the first step, the facial image of the user is obtained as the initial image. The frequency and time-domain features of the user's voice are extracted simultaneously. The voice features are used to encrypt the initial image, and the encrypted image is saved as the original comparison image. (2) In the second step, dual authentication of speech recognition and facial recognition is performed. If it is necessary to verify the identity of the user after a certain period of time, the latest voice data and face image data of the user must be extracted for the second time. The extracted second speech frequency-domain and time-domain features are used to decrypt the original encrypted contrast image. The voice authentication is passed if the decryption is successful. At this point, the decrypted initial face image is used to perform face recognition on the face image extracted for the second time. If the recognition is successful, the face authentication is passed. At this time, the dual speech and facial recognition authentication is passed, and subsequently it is judged that the user is successfully authenticated. (3) The third step consists of encrypting and saving the facial image that passed the second authentication based on the speech frequency and time-domain features extracted the second time. If a third authentication is required, the image encrypted for the second time will be used as the latest comparison image. The first and second steps are repeated to obtain the certification result.

When associating the speech feature information with the chaotic encryption algorithm, it is necessary to design the parameter mapping relationship in order to meet the operational needs of Equation (5). Using the solution presented in this article, developers can design their own chaotic parameter mapping relationship and input voice features information into the algorithm. For example, when selecting the peak number of a single waveform, the following can be varied: single-syllable orders; the value range of logistic mapping parameter μ; initial chaotic mapping values; and the output index values of different chaotic output vectors. The changes in these mapping relations can significantly impact the encryption result of the image without affecting the sensitivity of the input data to decryption. In addition, when selecting the voice features, other frequency or time-domain voice features can be chosen, such as the discrete cosine transform (DCT) or discrete wavelet transform (DWT) used in the frequency-domain feature analysis. The DCT transformation is simple and practical and can be calculated quickly. The DWT transformation is a multi-resolution analysis method that increases the concealment of the algorithm. The time-domain features can use short-term average energy, short-term average amplitude, short-term average zero-crossing rate and short-term autocorrelation analysis. Other chaotic maps can also be used in the encryption algorithm, such as singer map, sine map, tent map, Chebyshev map, circle map, cubic map, sinusoidal map, etc. These methods can be used to expand and design a unique authentication scheme [20–22].

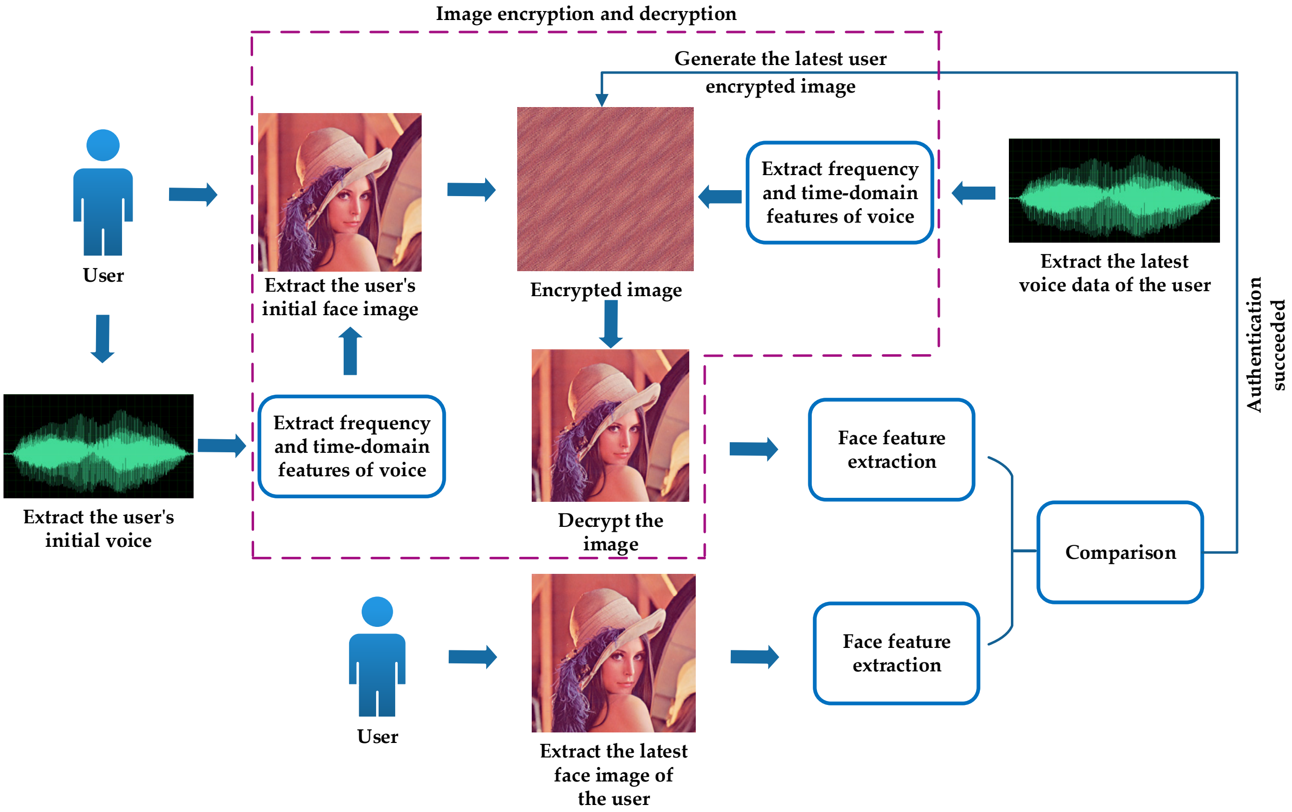


Image encryption and decryption
Generate the latest user encrypted image
User
Extract the user's initial face image
Encrypted image
Extract frequency and time-domain features of voice
Extract the latest voice data of the user
Authentication succeeded
Extract the user's initial voice
Extract frequency and time-domain features of voice
Decrypt the image
Face feature extraction
Comparison
Face feature extraction
User
Extract the latest face image of the user

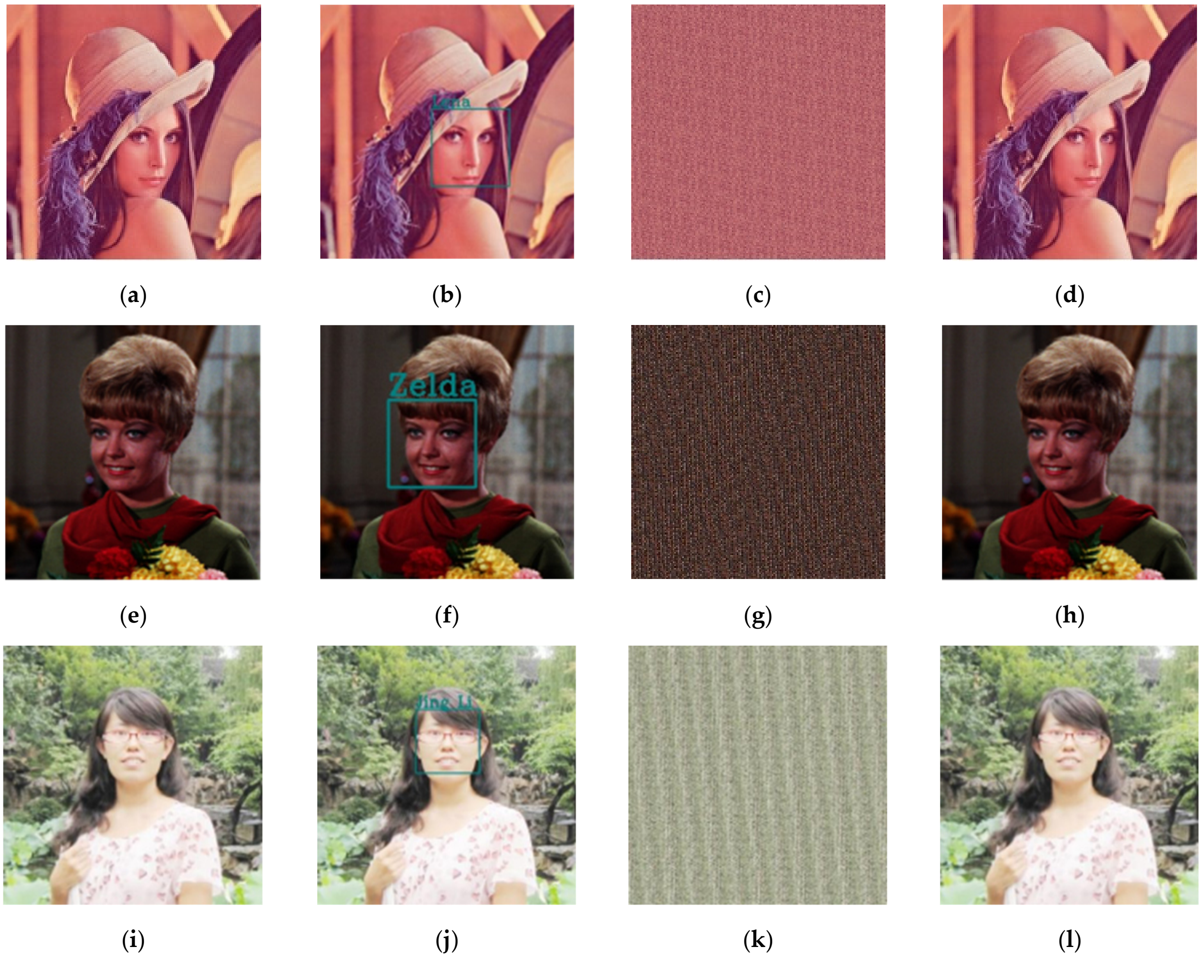
Figure 5. Authentication process of the identification method in this article.

4. Experimental Results and Security Analysis

4.1. Experimental Results

In the experiment, Python was used for facial recognition, and Matlab for voice features extraction and image encryption and decryption. The Haar classification detector and LBPH recognizer of the OpenCV library were used for facial recognition. The initial value of a chaotic map should be between zero and one. Therefore, it was selected as 1/10th of the number of single-syllable and single-waveform peaks. The logistic map parameter μ was selected as 3.85 plus 1/100th of the number of single-syllable and single-waveform peaks. This ensures the data generated during the iterative process are in the chaotic region. The number of iterations of the scrambling formula was the sum of the number of rows and the number of columns of an image. The number of iterations of the chaotic map was chosen as 3000, and the index values of L_{nL1} and L_{nL2} were determined by Equation (4). An integer above 400 was used as the index value for the selection of the chaotic data vector, so that it could contain enough chaotic output data. The experiments were performed on

80 groups of face images and voice data, all of which passed the encryption and decryption tests. Figure 6 shows the effects of image recognition, encryption and decryption.



(**a**) (**b**) (**c**) (**d**)

(**e**) (**f**) (**g**) (**h**)

(**i**) (**j**) (**k**) (**l**)

Figure 6. Image recognition, encryption and decryption. (**a**,**e**,**i**) Target image; (**b**,**f**,**j**) Face recognition image; (**c**,**g**,**k**) Logistic chaotic encryption; (**d**,**h**,**l**) Decrypted image.

Table 1 shows the accuracy rate of the method proposed in this paper compared with other schemes reported in the literature.

Table 1. Comparison of recognition results.

	Scheme	Year	Accuracy Rate (%)
Ref. [2]	Iris Recognition	2001	98.4
Ref. [3]	Retina Identification	2011	99.75
Ref. [4]	Face Authentication	2019	96
Ref. [6]	Signature Recognition	2020	97.5
Ref. [7]	Finger vein Recognition	2020	99.56
Ref. [17]	Face Verification + ID Image Recognition	2019	97.5
Ref. [18]	Facial Authentication + Speaker Recognition	2021	82.71
Proposed (Tolerance 100 Hz)	Facial Authentication + Voice Recognition	2022	100
Proposed (Tolerance 60 Hz)	Facial Authentication + Voice Recognition	2022	91.3
Proposed (Tolerance 30 Hz)	Facial Authentication + Voice Recognition	2022	61.3
Proposed (Tolerance 10 Hz)	Facial Authentication + Voice Recognition	2022	23.8

4.2. Security Analysis

The identity recognition method used in this article utilizes encryption and decryption methods and compares and authenticates both voice and image. No identity authentication scheme similar to the one proposed in this article is found in the various existing references. Other identity authentication schemes only collect a single biometric feature of the user for identity authentication, such as fingerprint authentication, face recognition, etc. Therefore, this article only analyzes the key sensitivity, the correlation between adjacent pixels, the resistance to differential attack and the size of the key space from the perspective of encryption.

4.2.1. Key Sensitivity Analysis

The applied system is a completely closed system and the key image is not visible from the outside. Therefore, the key sensitivity is the most important item for verifying the scheme [23–25]. Lena image and various voice data samples are used in the scheme verification. When comparing images, the normalized cross-correlation parameter C_{ncc} shown in Equation (7) is used in addition to the visual observation to measure the changes in the image.

$$C_{ncc} = \frac{\sum_{x_c=0}^{N-1}\sum_{y_c=0}^{N-1} g_c(x_c, y_c) f_c(x_c, y_c)}{\sum_{x_c=0}^{N-1}\sum_{y_c=0}^{N-1} f_c{}^2(x_c, y_c)} \tag{7}$$

In Equation (7), f_c and g_c are the two images to be compared, and x_c and y_c are the index values of the image matrix. The experimental data shown in Tables 2–7 were obtained after various experimental comparisons. In the tables, F_p is the peak frequency of the speech spectrum; n_w1 to n_w6 are the number of single-waveform peaks of six single syllables; and C_{nccr}, C_{nccg} and C_{nccb} are the normalized cross-correlation parameters of the red, green and blue channels of the image.

Table 2. The normalized cross-correlation when fine-tuning the number of crests.

F_p (Hz)	n_w1~n_w6	$\lvert C_{nccr} - 1 \rvert_{min}$	$\lvert C_{nccg} - 1 \rvert_{min}$	$\lvert C_{nccb} - 1 \rvert_{min}$
55	5, 4, 2, 3, 2, 4	0.0675	0.4152	0.5422
650	5, 3, 5, 2, 5, 3	0.0625	0.4150	0.5421
1220	3, 2, 4, 3, 5, 2	0.0682	0.4141	0.5472

Table 3. The normalized cross-correlation when fine-tuning the peak frequency.

F_p (Hz)	n_w1~n_w6	$\lvert C_{nccr} - 1 \rvert_{min}$	$\lvert C_{nccg} - 1 \rvert_{min}$	$\lvert C_{nccb} - 1 \rvert_{min}$
30~35	4, 3, 5, 2, 3, 2	0.0689	0.4138	0.5469
660~665	3, 5, 2, 4, 2, 3	0.0685	0.4156	0.5479
1220~1205	2, 3, 5, 3, 4, 2	0.0689	0.4152	0.5479

Table 4. The correlation coefficients between adjacent pixels in the plain and cipher images.

Fp (Hz)	n_w1~n_w6	Horizontal			Vertical			Diagonal		
		R	G	B	R	G	B	R	G	B
55	5, 4, 2, 3, 2, 4	−0.0772	−0.0213	−0.0028	0.1237	0.0857	0.1114	−0.0389	0.0517	0.0770
1200	2, 3, 5, 3, 4, 2	0.1764	0.1109	0.0494	−0.0536	−0.0542	−0.0432	0.0043	0.0324	0.0037
Original image		0.9798	0.9690	0.9329	0.9894	0.9824	0.9578	0.9696	0.9552	0.9181

Table 2 shows the experimental results of the normalized cross-correlation calculated by adding one to the number of single-syllable waveforms of the voice key without changing the peak frequency. Table 3 shows the experimental results of the normalized cross-correlation calculated after the number of single waveforms remained unchanged and the peak frequency was fine-tuned. The parameters C_{nccr} (red channel), C_{nccg} (green channel) and C_{nccb} (blue channel) are normalized cross-correlations between the encrypted image before fine-tuning and the encrypted image after fine-tuning of a parameter. It can be observed from the experimental data provided in these two tables that when the encrypted input data undergo a small change, the encrypted image changes significantly in the green and blue channels. As the overall tone of the image is reddish, the red channel changes slightly. Figures 6 and 7 are used to intuitively illustrate this remarkable change.

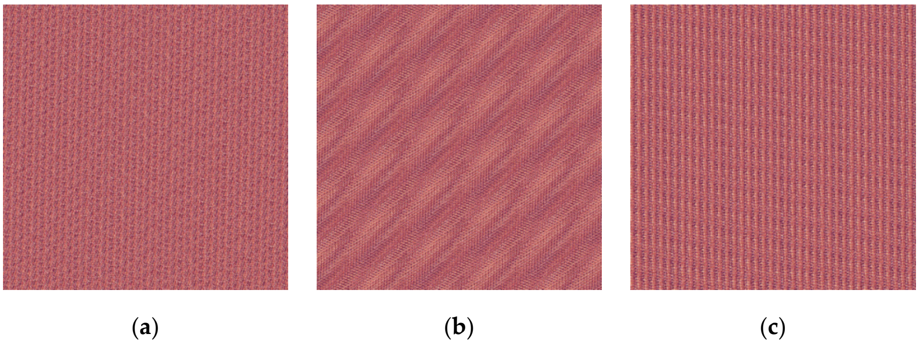


Figure 7. Speech peak frequency is unchanged, and the waveform number is fine-tuned for image decryption. (**a**) Encrypted image when n_w1~n_w6 are 5, 3, 5, 2, 5, 3; (**b**) Encrypted image when n_w1~n_w6 are 5, 4, 5, 2, 5, 3; (**c**) Decrypted image that fails after fine-tuning the parameters.

Figure 7 shows that when F_p is 650 Hz, n_w1~n_w6 are set to 5, 4, 5, 2, 5, 3 to decrypt the encrypted image of n_w1~n_w6 (5, 3, 5, 2, 5, 3). Figure 8 shows that when n_w1~n_w6 are 3, 5, 2, 4, 2, 3, the encrypted image with F_p of 662 Hz is used to decrypt the encrypted image with F_p of 661 Hz. From the experimental data samples presented in the table and the experimental images shown in Figures 7 and 8, it can be observed that after the fine-tuning of the number of crests and peak frequency, the image decryption cannot be performed normally, and it is very sensitive to the input parameters. The decrypted image is completely unrecognizable as long as there is a slight input change, indicating that the chaotic characteristics of the encryption algorithm have been included in the encryption process.

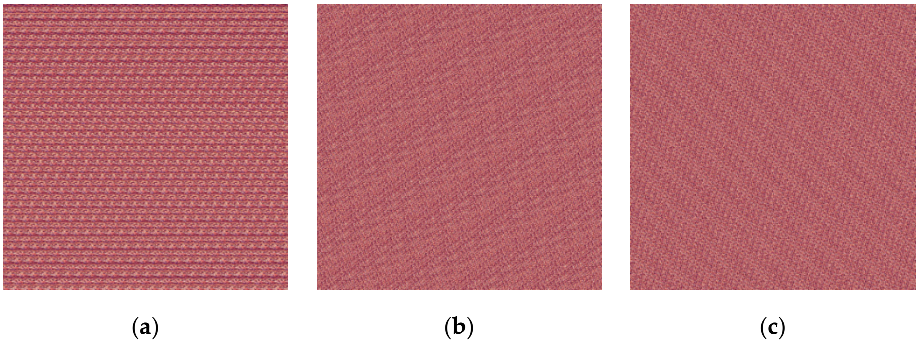


Figure 8. Unchanged n_w1~n_w6, image decryption with fine-tuned voice peak frequency. (**a**) Encrypted image when F_p is equal to 661 Hz; (**b**) Encrypted image when F_p is equal to 662 Hz; (**c**) Decrypted image that fails after fine-tuning the parameters.

It can be concluded from the experimental process that the algorithm is suitable for remote identification. The authentication scheme is a completely closed system where the encrypted image information is completely invisible to the outside. Therefore, it does not involve the anti-attack and crack verification of the encrypted information and only the sensitivity of the algorithm should be verified. It can be gathered through experimental verification and data analysis that the scheme is feasible. The image decoding is very sensitive to the input information. If the features of the input voice information are slightly different from those of the voice information input provided last time, the decoded image cannot be recognized at all.

This sensitivity has advantages and disadvantages. On the one hand, it effectively satisfies the system operation requirements and significantly improves the safety and resolution of the system. On the other hand, when the verifier performs information authentication every other year, there can be a slight change in the voice characteristics. This kind of change is mainly related to the small feature changes in the audio domain, and even a 1 Hz input difference will cause the authentication to fail. As long as the verifier's single-syllable voice remains unchanged, the number of single peaks of a single waveform in the time domain will not change. However, a person with the same voice frequency characteristics may still have very small changes between two authentications. This shortcoming can be solved by setting the tolerance of the peak frequency during the programming process and decoding the encrypted image within the tolerance range. This overcomes the misjudgment caused by the high sensitivity of the algorithm.

4.2.2. Correlation between Adjacent Pixels

The correlation between adjacent pixels is strong for plaintext images, and it can be easily cracked by the attackers. Therefore, in order to increase the difficulty of decryption and improve the security for ciphertext images, the correlation between adjacent pixels should be reduced [26,27]. In this article, the horizontal, vertical and diagonal adjacent pixels of the plaintext and ciphertext images of Lena from red, green and blue channels are analyzed for correlation under different *Fp* conditions. The correlation calculation formula is as follows:

$$\begin{cases} r_{xy} = \frac{cov(x,y)}{\sqrt{D(x)}\sqrt{D(y)}} \\ cov(x,y) = \frac{1}{N}\sum\limits_{i=1}^{N}(x_i - E(x))(y_i - E(y)) \\ E(x) = \frac{1}{N}\sum\limits_{i=1}^{N} x_i \\ E(y) = \frac{1}{N}\sum\limits_{i=1}^{N} y_i \\ D(x) = \frac{1}{N}\sum\limits_{i=1}^{N}(x_i - E(x))^2 \\ D(y) = \frac{1}{N}\sum\limits_{i=1}^{N}(y_i - E(y))^2 \end{cases} \tag{8}$$

where, x_i and y_i (i = 1, 2, . . . , N) are the brightness values of two adjacent pixels from each color channel, and r_{xy} is the correlation coefficient of two adjacent pixels. The size of the pixel sample set selected in the image is denoted by N, where all pixels are selected. The expectation and variance of the variable x are denoted by $E(x)$ and $D(x)$, respectively, while $E(y)$ and $D(y)$ denote the expectation and variance of variable y, respectively. Table 4 shows the coefficients of two adjacent pixels of the encrypted ciphertext image under different *Fp* conditions, different color channels and different directions. These coefficients are significantly reduced and are all close to zero. Table 5 shows a comparison with other algorithms.

Table 5. The comparison of image correlation coefficients.

Image	Horizontal			Vertical			Diagonal		
	Ref. [28]	**Ref. [29]**	**Ref. [30]**	**Ref. [28]**	**Ref. [29]**	**Ref. [30]**	**Ref. [28]**	**Ref. [29]**	**Ref. [30]**
Lena	0.01534	0.0214	−0.038118	0.01391	0.0465	−0.029142	0.01399	0.009	0.002736

4.2.3. Ability of Resisting Differential Attack

The image encryption algorithm should effectively resist differential attacks. Therefore, it should be very sensitive to subtle changes in the original image. The number of changing pixel rate (NPCR) and the unified averaged changed intensity (UACI) are the two most common quantitative methods used to evaluate the robustness of an image encryption algorithm against differential attacks. The ideal values of NPCR and UACI are 99.6094% and 33.4635% [31–34], respectively. These two metrics are defined by the following expressions:

$$NPCR(C_1, C_2) = \sum_{j=1}^{j=N} \sum_{i=1}^{i=M} \frac{D(i,j)}{M \times N} \times 100\% \tag{9}$$

$$UACI = (C_1, C_2) = \sum_{j=1}^{j=N} \sum_{i=1}^{i=M} \frac{|c_1(i,j) - c_2(i,j)|}{M \times N \times 255} \times 100\% \tag{10}$$

$$D(i,j) = \begin{cases} 0, C_1(i,j) = C_2(i,j) \\ 1, C_1(i,j) \neq C_2(i,j) \end{cases} \tag{11}$$

where, C_1 and C_2 represent two different encrypted images, and their corresponding plaintext images only differ in pixel values at one position. Let $C_1(i, j)$ and $C_2(i, j)$ be the ith row and jth column pixels of two cipher images C_1 and C_2, respectively. The size of the image is $M \times N$ and $D(i, j)$ represents whether the corresponding pixel values $C_1(i, j)$ and $C_2(i, j)$ of the two images are equal or not. Table 6 shows the NPCR and UACI values of the red, green and blue channels of Lena image for varying values of n_w1~n_w6 while keeping Fp as constant, and varying values of Fp while keeping n_w1~n_w6 constant. Table 7 shows the comparison with other algorithms. The experiments show that the NPCR value of Lena ciphertext image fine-tuned by Fp or n_w1~n_w6 is very close to the ideal value of 99.6094%.

Table 6. The NPCR and UACI for the cipher images.

Fp (Hz)	n_w1~n_w6	R		G		B	
		NPCR	UACI	NPCR	UACI	NPCR	UACI
55 55	5, 4, 2, 3, 2, 4 5, 4, 3, 3, 2, 4	0.9938	0.2271	0.9951	0.2411	0.9916	0.1521
55 56	5, 4, 2, 3, 2, 4 5, 4, 2, 3, 2, 4	0.9920	0.2139	0.9946	0.2377	0.9908	0.1497
1200 1200	3, 2, 4, 3, 5, 2 3, 2, 5, 3, 5, 2	0.9893	0.2105	0.9940	0.2287	0.9891	0.1376
1200 1201	3, 2, 4, 3, 5, 2 3, 2, 4, 3, 5, 2	0.9925	0.2138	0.9943	0.2375	0.9909	0.1496

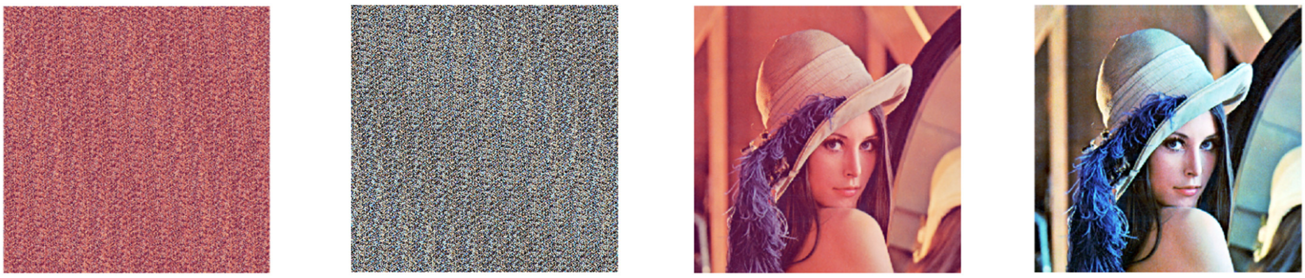
Table 7. The comparison results of NPCR and UACI.

	Ref. [35]	Ref. [36]	Ref. [37]	Ref. [38]	Ref. [39]
NPCR (%)	99.54	99.6146	99.6048	99.4602	99.655
UACI (%)	28.27	33.5113	33.2966	33.2161	33.516

It can be seen from the experimental data that there is a certain gap between the UACI value obtained using the proposed method and the values provided in the existing literature. This is caused by the uneven gray-scale distribution of the experimental images. Histogram equalization could be added to the experiment to optimize the UACI. Table 8 shows the optimized UACI, which is close to the ideal value and is equivalent to the UACI values given in other references. Figure 9 shows the optimized images and Figure 10 depicts the histogram of the original image and the decrypted image before and after equalization.

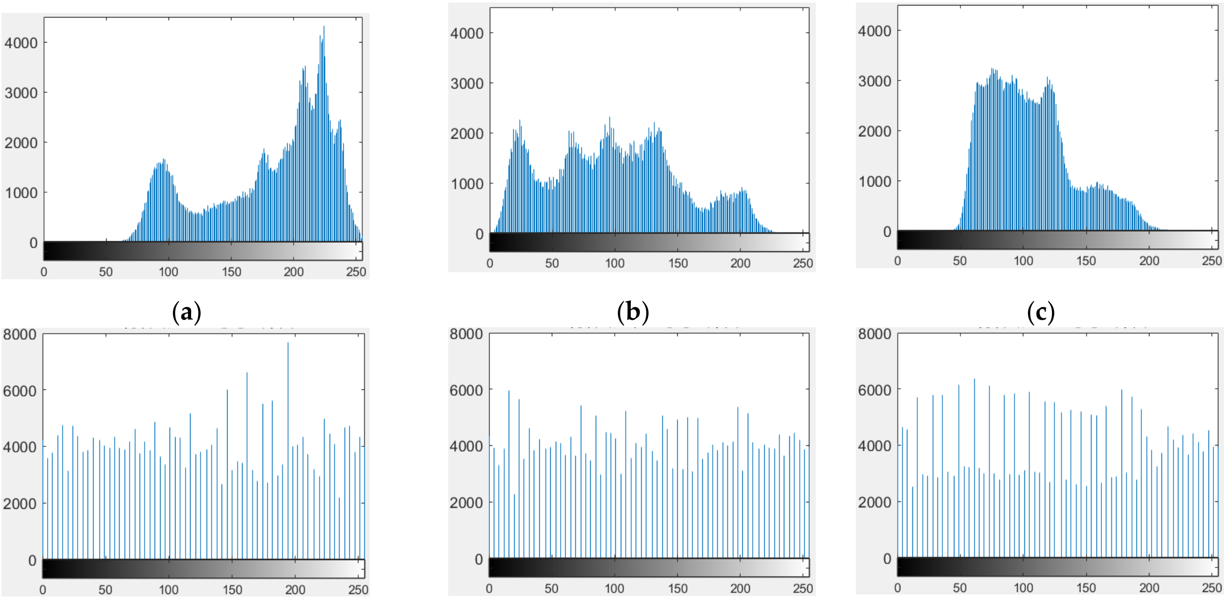
Table 8. The optimized UACI.

Fp **(Hz)**	n_w**1~**n_w**6**	**R** **UACI**	**G** **UACI**	**B** **UACI**
55 55	5, 4, 2, 3, 2, 4 5, 4, 3, 3, 2, 4	0.3559	0.3422	0.3417
55 56	5, 4, 2, 3, 2, 4 5, 4, 2, 3, 2, 4	0.3383	0.3394	0.3387
1200 1200	3, 2, 4, 3, 5, 2 3, 2, 5, 3, 5, 2	0.3336	0.3284	0.3177
1200 1201	3, 2, 4, 3, 5, 2 3, 2, 4, 3, 5, 2	0.3380	0.3392	0.3385



(**a**) (**b**) (**c**) (**d**)

Figure 9. The optimized renderings. (**a**) Encrypted image before optimization; (**b**) Optimized encrypted image; (**c**) Decrypted image without optimization; (**d**) Optimized decrypted image.



(**a**) (**b**) (**c**)

(**d**) (**e**) (**f**)

Figure 10. Histograms of original and encrypted images. (**a**) Histogram of the red channel of the original image before equalization; (**b**) Histogram of the green channel of the original image before equalization; (**c**) Histogram of the blue channel of the original image before equalization; (**d**) Histogram of the red channel of the decrypted image after equalization; (**e**) Histogram of the green channel of the decrypted image after equalization; (**f**) Histogram of the blue channel of the decrypted image after equalization.

4.2.4. Key Space Analysis

The size of the key space directly influences the quality of the encryption algorithm, and also determines the probability of its being successfully cracked. Generally, an encryption algorithm with a small key space is likely to be cracked by brute force without the use of any particular decryption skills. Research shows that the key space of the algorithm should be greater than 2^{100} to prevent the possibility of being cracked [40–42].

When counting the key space, a 6-digit voice password is considered. The value range of f_m is 50–2000 Hz and the value interval is 1 Hz. The value range of $k1$, $k2$, $k3$ and $k6$ is 20–1000 and the value interval is 10^{-7}. The value range of $k7$ and $k8$ is 10^3–10^{10}, the value interval is 1, and the calculated key space is 2^{232}. If the number of voice cipher bits is increased, the capacity of the key space can be further increased.

In addition, the commonly used analytical methods of encryption algorithms also include histogram and information entropy analysis [43–46]. The encryption method proposed in this article is used for face recognition. Therefore, the gray distribution of the image can be arbitrarily changed while keeping the facial characteristics unchanged to construct the required histogram distribution and information entropy value without affecting the application of the algorithm.

5. Conclusions

This paper proposed an image encryption and decryption algorithm based on a voice data key. The remote identification scheme designed using this algorithm improves the robustness of the traditional face recognition verification method. The experiments show that the encryption and decryption algorithms are highly sensitive to the key and the sensitivity is adjustable. If the time-domain waveform of the user's voice differed by a single peak, or the peak frequency differs by 1 Hz, the encrypted image cannot be decrypted. The algorithm key space 2^{232} has sufficient resistance to exhaustive attacks. On the basis of this algorithm, the key space can be further increased by adjusting the parameters. The correlation experiment between adjacent pixels showed that the maximum value of correlation coefficients is 0.1764, which appeared in the horizontal direction of the encrypted red channel at a peak frequency of 1200 Hz. The analysis of the differential attack experiment revealed that the minimum value of NPCR was 0.9891, which appeared in the encrypted blue channel at the peak frequency of 1200 Hz, and the UACI value was also close to the ideal value. The experiments showed that this scheme is feasible and has strong scalability. It can be extended by selecting different voice features of the users, by selecting different chaotic map, and by selecting different parameters of Equation (4) (k7, k8 and k9).

The highlights of this scheme are as follows. First, it uses two kinds of feature information of the user for authentication. Second, it uses the voice feature as a key. Third, it is suitable for large-scale promotion and application, i.e., the client only needs a computer or a mobile phone as an input device, and no other hardware devices are required to perform remote authentication from different places. The limitation of this scheme is that the identity authentication cannot be performed for deaf and mute individuals. One possible solution is that deaf and mute individuals may manually enter a key, but this will reduce the security of the system. In addition, the scheme can currently only authenticate one targeted individual, and only people can be authenticated. If the scheme is applied to other goals, the key extraction problem and the image recognition problem should also be solved. The next step would be to integrate artificial intelligence into the system to solve the problem of image recognition for authentication of different targets.

Author Contributions: Conceptualization, J.L. and T.F.; methodology, J.L., T.F. and C.F; software, T.F.; validation, J.L.; writing—original draft preparation, J.L., L.H. and T.F.; writing—review and editing, J.L., C.F. and L.H. 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 (51774092), the Daqing City Guiding Science and Technology Plan Project (zd-2021-87), National Natural Science

Foundation of Heilongjiang Province (LH2020E012) and the China Postdoctoral Science Foundation (2016M601399).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

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

References

1. How, K.W.; Han, P.Y.; Yin, O.S.; Yen, Y.H. Spatiotemporal spectral histogramming analysis in hand gesture signature recognition. *J. Intell. Fuzzy Syst.* **2021**, *40*, 4275–4286. [CrossRef]
2. Lim, S.L.; Lee, K.L.; Byeon, O.B.; Kim, T.K. Efficient Iris Recognition through Improvement of Feature Vector and Classifier. *ETRI J.* **2001**, *23*, 61–70. [CrossRef]
3. Barkhoda, W.; Akhlaqian, F.; Amiri, M.D.; Nouroozzadeh, M.S. Retina identification based on the pattern of blood vessels using fuzzy logic. *EURASIP J. Adv. Signal Process.* **2011**, *2011*, 113. [CrossRef]
4. Parvathi, R.; Sankar, M. An Exhaustive Multi Factor Face Authentication Using Neuro-Fuzzy Approach. *Wirel. Pers. Commun.* **2019**, *109*, 2353–2375. [CrossRef]
5. Park, H.; Kim, T. User Authentication Method via Speaker Recognition and Speech Synthesis Detection. *Secur. Commun. Netw.* **2022**, *2022*, 5755785. [CrossRef]
6. Lakkannavar, B.F.; Kodabagi, M.M.; Naik, S.P. Signature Recognition and Verification Using Zonewise Statistical Features. In *International Conference on Computer Networks, Big Data and IoT*; Springer: Cham, Switzerland, 2020; pp. 748–757. [CrossRef]
7. Boucherit, I.; Zmirli, M.O.; Hentabli, H.; Rosdi, B.A. Finger vein identification using deeply-fused Convolutional Neural Network. *J. King Saud Univ.-Comput. Inf. Sci.* **2020**, *34*, 646–656. [CrossRef]
8. Sodhro, A.H.; Sennersten, C.; Ahmad, A. Towards Cognitive Authentication for Smart Healthcare Applications. *Sensors* **2022**, *22*, 2101. [CrossRef]
9. Khan, A.; Geng, S.; Zhao, X.; Shah, Z.; Jan, M.U.; Abdelbaky, M.A. Design of MIMO antenna with an enhanced isolation technique. *Electronics* **2020**, *9*, 1217. [CrossRef]
10. Noor, K.; Jan, T.; Basheri, M.; Ali, A.; Khalil, R.A.; Zafar, M.H.; Ashraf, M.; Babar, M.I.; Shah, S.W. Performances Enhancement of Fingerprint Recognition System Using Classifiers. *IEEE Access* **2018**, *7*, 5760–5768. [CrossRef]
11. Gil Hong, H.; Lee, M.B.; Park, K.R. Convolutional Neural Network-Based Finger-Vein Recognition Using NIR Image Sensors. *Sensors* **2017**, *17*, 1297. [CrossRef]
12. AlMahafzah, H.; Alrwashdeh, M.Z. A survey of multi- biometric systems. *Int. J. Comput. Appl.* **2012**, *43*, 36–43. [CrossRef]
13. Gad, R.; Elsayed, A.; Zorkany, M.; Elfishawy, N. Multi-biometric systems: A state of the art survey and research directions. *Int. J. Adv. Comput. Sci. Appl.* **2015**, *6*, 128–138. [CrossRef]
14. Dehghani, A.; Ghassabi, Z.; Moghddam, H.A.; Moin, M.-S. Human recognition based on retinal images and using new similarity function. *EURASIP J. Image Video Process.* **2013**, *2013*, 58. [CrossRef]
15. Hou, B.; Yan, R. Convolutional Autoencoder Model for Finger-Vein Verification. *IEEE Trans. Instrum. Meas.* **2019**, *69*, 2067–2074. [CrossRef]
16. Fernandez, F.A. Biometric Sample Quality and Its Application to Multimodal Authentication Systems. Ph.D. Thesis, Universidad Politecnica de Madrid (UPM), Madrid, Spain, 2008.
17. Wu, X.; Xu, J.; Wang, J.; Li, Y.; Li, W.; Guo, Y. Identity authentication on mobile devices using face verification and ID image recognition. *Procedia Comput. Sci.* **2019**, *162*, 932–939. [CrossRef]
18. Saxena, N.; Varshney, D. Smart Home Security Solutions using Facial Authentication and Speaker Recognition through Artificial Neural Networks. *Int. J. Cogn. Comput. Eng.* **2021**, *2*, 154–164. [CrossRef]
19. Han, C.Y. An image encryption algorithm based on modified logistic chaotic map. *Optik* **2019**, *181*, 779–785. [CrossRef]
20. Pak, C.; An, K.; Jang, P.; Kim, J.; Kim, S. A novel bit-level color image encryption using improved 1D chaotic map. *Multimedia Tools Appl.* **2018**, *78*, 12027–12042. [CrossRef]
21. Parvaz, R.; Zarebnia, M. A combination chaotic system and application in color image encryption. *Opt. Laser Technol.* **2018**, *101*, 30–41. [CrossRef]
22. Li, C.; Luo, G.; Qin, K.; Li, C. An image encryption scheme based on chaotic tent map. *Nonlinear Dyn.* **2016**, *87*, 127–133. [CrossRef]
23. Zhou, Y.; Bao, L.; Chen, C.L.P. Image encryption using a new parametric switching chaotic system. *Signal Process.* **2013**, *93*, 3039–3052. [CrossRef]
24. Muñoz-Guillermo, G. Image encryption using q-deformed logistic map. *Inf. Sci.* **2021**, *552*, 352–364. [CrossRef]
25. Li, R.; Liu, Q.; Liu, L. Novel image encryption algorithm based on improved logistic map. *IET Image Process.* **2019**, *13*, 125–134. [CrossRef]
26. Liu, W.; Sun, K.; Zhu, C. A fast image encryption algorithm based on chaotic map. *Opt. Lasers Eng.* **2016**, *84*, 26–36. [CrossRef]

27. Wang, X.; Chen, S.; Zhang, Y. A chaotic image encryption algorithm based on random dynamic mixing. *Opt. Laser Technol.* **2021**, *138*, 106837. [CrossRef]
28. Shakiba, A. A novel randomized one-dimensional chaotic Chebyshev mapping for chosen plaintext attack secure image encryption with a novel chaotic breadth first traversal. *Multimed. Tools Appl.* **2019**, *78*, 34773–34799. [CrossRef]
29. Zhen, P.; Zhao, G.; Min, L.; Jin, X. Chaos-based image encryption scheme combining DNA coding and entropy. *Multimed. Tools Appl.* **2016**, *75*, 6303–6319. [CrossRef]
30. Zhou, M.; Wang, C. A novel image encryption scheme based on conservative hyperchaotic system and closed-loop diffusion between blocks. *Signal Process.* **2020**, *171*, 107484. [CrossRef]
31. Naim, M.; Pacha, A.A.; Serief, C. A novel satellite image encryption algorithm based on hyperchaotic systems and Josephus problem. *Adv. Space Res.* **2021**, *67*, 2077–2103. [CrossRef]
32. Hua, Z.Y.; Zhou, Y.C. Image encryption using 2D Logistic-adjusted-Sine map. *Inform. Sci.* **2016**, *339*, 237–253. [CrossRef]
33. Zhang, Y. Statistical test criteria for sensitivity indexes of image cryptosystems. *Inf. Sci.* **2021**, *550*, 313–328. [CrossRef]
34. Alghafis, A.; Munir, N.; Khan, M.; Hussain, I. An Encryption Scheme Based on Discrete Quantum Map and Continuous Chaotic System. *Int. J. Theor. Phys.* **2020**, *59*, 1227–1240. [CrossRef]
35. Huang, C.-K.; Liao, C.W.; Hsu, S.L.; Jeng, Y.C. Implementation of gray image encryption with pixel shuffling and gray-level encryption by single chaotic system. *Telecommun. Syst.* **2011**, *52*, 563–571. [CrossRef]
36. Wu, Y.; Zhou, Y.; Noonan, J.P.; Agaian, S. Design of image cipher using latin squares. *Inf. Sci.* **2013**, *264*, 317–339. [CrossRef]
37. Zhu, H.; Dai, L.; Liu, Y.; Wu, L. A three-dimensional bit-level image encryption algorithm with Rubik's cube method. *Math. Comput. Simul.* **2021**, *185*, 754–770. [CrossRef]
38. Bakhshandeh, A.; Eslami, Z. An authenticated image encryption scheme based on chaotic maps and memory cellular automata. *Opt. Lasers Eng.* **2013**, *51*, 665–673. [CrossRef]
39. Song, C.-Y.; Qiao, Y.-L.; Zhang, X.-Z. An image encryption scheme based on new spatiotemporal chaos. *Optik* **2013**, *124*, 3329–3334. [CrossRef]
40. Li, T.; Du, B.; Liang, X. Image Encryption Algorithm Based on Logistic and Two-Dimensional Lorenz. *IEEE Access* **2020**, *8*, 13792–13805. [CrossRef]
41. Wu, Y.; Zhang, L.; Qian, T.; Liu, X.; Xie, Q. Content-adaptive image encryption with partial unwinding decomposition. *Signal Process.* **2020**, *181*, 107911. [CrossRef]
42. Shen, Y.; Tang, C.; Xu, M.; Lei, Z. Optical selective encryption based on the FRFCM algorithm and face biometric for the medical image. *Opt. Laser Technol.* **2021**, *138*, 106911. [CrossRef]
43. Li, Z.; Peng, C.; Tan, W.; Li, L. A Novel Chaos-Based Image Encryption Scheme by Using Randomly DNA Encode and Plaintext Related Permutation. *Appl. Sci.* **2020**, *10*, 7469. [CrossRef]
44. Lin, C.-H.; Hu, G.-H.; Chan, C.-Y.; Yan, J.-J. Chaos-Based Synchronized Dynamic Keys and Their Application to Image Encryption with an Improved AES Algorithm. *Appl. Sci.* **2021**, *11*, 1329. [CrossRef]
45. Wei, T.; Lin, P.; Zhu, Q.; Yao, Q. Instability of impulsive stochastic systems with application to image encryption. *Appl. Math. Comput.* **2021**, *402*, 126098. [CrossRef]
46. Jiang, X.; Xiao, Y.; Xie, Y.; Liu, B.; Ye, Y.; Song, T.; Chai, J.; Liu, Y. Exploiting optical chaos for double images encryption with compressive sensing and double random phase encoding. *Opt. Commun.* **2020**, *484*, 126683. [CrossRef]