mathematics

MDPI

Article

Security Analysis and Improvement of Dual Watermarking Framework for Multimedia Privacy Protection and Content Authentication

Ming Li [1,2,*] and Yange Yue [1]

[1] College of Computer and Information Engineering, Henan Normal University, Xinxiang 453007, China
[2] Key Laboratory of Artificial Intelligence and Personalized Learning in Education of Henan Province, Xinxiang 453007, China
* Correspondence: liming@htu.edu.cn

Abstract: The demand for using multimedia network infrastructure for transmission grows with each passing day. Research scholars continue to develop new algorithms to strengthen the existing network security framework in order to ensure the privacy protection and content authentication of multimedia content and avoid causing huge economic losses. A new technology for multimedia image copyright protection and content authentication has been proposed. The innovations lie in the use of an inter-block coefficient difference algorithm to embed robust watermarking in the transform domain, and the same fragile watermark is embedded twice in the spatial domain so that any tiny tampering can be identified and located. A new encryption algorithm combined with Arnold transform is used to encrypt data before embedding. However, some security vulnerabilities were found, and successful cryptanalysis and attack were conducted. Subsequently, an improved scheme was proposed to improve the security and tamper detection ability of the original watermarking scheme and recover the tampered robust watermark. The results show that the improved scheme is safer and more reliable and shows good performance in tampering detection and the recovery robustness of the watermark.

Keywords: image authentication; dual blind watermark; tamper detection; safety analysis; tamper recovery

MSC: 68U10

Citation: Li, M.; Yue, Y. Security Analysis and Improvement of Dual Watermarking Framework for Multimedia Privacy Protection and Content Authentication. *Mathematics* **2023**, *11*, 1689. https://doi.org/10.3390/math11071689

Academic Editor: Jonathan Blackledge

Received: 28 February 2023
Revised: 28 March 2023
Accepted: 29 March 2023
Published: 1 April 2023

1. Introduction

With the popularity of the mobile internet and the continuous evolution of emerging information technologies such as 5G, cloud computing, AI, etc., network development has entered a new era of digital intelligent media. People have generated vast amounts of information resources through the explosive growth of network equipment around them, resulting in a huge exchange of multimedia content on the Internet, including secret information intended to be secretly transmitted by some individuals or organizations [1–3]. However, the openness and convenience of the Internet as a public network make the protection of multimedia data content face unprecedented challenges. With the emergence of powerful devices and easy-to-use software, the threats to copyright protection and authentication digital content are also increasing [4–6].

Digital watermarking technology is a common technical means to solve these problems. Researchers around the world have proposed a variety of watermarking schemes for privacy protection and content authentication [7–11]. Robust watermarking and fragile watermarking are the two most commonly used watermarking schemes. Robust watermarking is to embed digital content that needs to be protected, such as copyright identification, into the carrier image as an invisible watermark signal, which can be extracted from the

distorted image even if attacked and is not easy to be modified or even deleted [12,13]. Fragile watermarking is to make the invisible watermark signal embedded in the carrier image as sensitive to any modification suffered by the host as possible so as to verify the integrity of the protected content, detect tampering and locate tampering [14]. However, a single digital watermark generally has only a single function of copyright protection or content authentication, while the actual needs of users are complex and changeable. Therefore, the double watermark scheme has attracted more and more attention [15–19].

In the face of the booming digital watermarking technology, researchers have listed the attack effect on the watermarking algorithm as an important standard to evaluate the security of the watermark [20,21], and some researchers have found some insecure watermarking schemes. Teng et al. [22] and Marco et al. [23] have successively questioned the security of a fragile watermarking scheme for image tampering detection based on a chaotic system proposed by Rawat et al. [24]. Teng et al. [22] proved that the scheme has a security vulnerability in that the embedded watermark information is easy to be extracted and replaced. Then, cryptanalysis and modification attacks were carried out and improvement measures were proposed to improve the security. Marco et al. [23] directly proved that the scheme could not be used to detect and locate the tampered region with bypassing the watermark verification program to tamper with the watermark information and proposed their own improved scheme. Moreover, based on the security vulnerability, embedded watermarks are easily replaced by attackers. Li et al. [25] carried out a replacement attack on a watermarking algorithm for remote sensing image copyright protection [26] without detecting the modification of the original image. Nan et al. [27] carried out extraction and replacement attacks on the double-color image watermarking scheme proposed by Su et al. [28] to embed color watermark images in color carrier images and proposed more secure improvement measures that can resist this attack. In recent years, researchers have also studied more and more image watermark tampering detection techniques [29–33], which not only play a basic role in copyright protection and content authentication but also enrich the functions of tampering detection and tampering recovery. Machine learning-based technologies, such as support vector machines, decision trees and naive Bayes, are also used in tampering detection methods. Niyishaka et al. [29] proposed a simple image mosaic forgery detection method, which uses a naive Bayesian model as the feature vector of classification to achieve the purpose of instantly distinguishing real images and forged images. However, image tampering detection algorithms are not all safe and reliable. Nandhini et al. [30] proposed a semi-fragile watermarking technology based on integer wavelet transform (IWT) and discrete cosine transform (DCT). The generated and embedded authentication watermarks are used for tampering detection and locating malicious attacks, while the generated recovered watermarks are used to create recovery labels to reconstruct the tampered watermark region. Oussama et al. [31] then analyzed the security of the scheme and found the existing security problems, carried out the watermark replacement attack on the image without being extracted the scheme alarm, and improved the watermark embedding coefficient and the encryption method of recovery label. Dadkhah et al. [32] proposed a watermark tamper detection and self-recovery algorithm based on singular value decomposition (SVD). The algorithm mainly generates different encrypted watermarks for different pixel blocks for tamper detection, and a random block mapping algorithm is used to extract the restored watermark bit from the least significant bit (LSB) of the mapping block. However, document [33] points out that the scheme has many shortcomings, such as easy access to private keys and easy discovery of mapping blocks, and cryptanalysis and modification attacks were carried out. The improved version enhances security by modifying the tamper detection and self-recovery watermark bit generation process. The existing watermarking attack technologies are only aimed at a single watermarking mechanism. As far as we know, there is no relevant literature and research on the attack of the dual watermarking mechanism.

With the application of dual watermarking technology in industry, medicine and other fields, there are more and more research schemes of the dual watermarking algorithm [18,34]. In

order to better meet the needs of users in various fields for the versatility of dual watermarking technology and to avoid major economic losses caused by some dual watermarking technologies with potential holes, it is of great significance to take security analysis and attack on the double watermark scheme. In this context, this paper analyzes the security of a dual watermark technology scheme for multimedia image copyright protection and content authentication proposed by Hurrah et al. [35]. The potential security vulnerabilities are found, and the robust watermark is destroyed, extracted and replaced successfully when the fragile watermark is not detected. In order to overcome the security problems existing in the scheme and realize the tamper detection and the recovery of tampered the robust watermark, this paper gives the improvement measures and adds a watermark recovery function to achieve accurate positioning and tamper recovery of tamper attacks. The results indicate that the improved scheme shows good performance in tampering detection and robust watermark recovery.

The main contributions of this paper are as follows:

(1) Conduct a security test on the scheme [35] and find potential security vulnerabilities.
(2) A cryptanalytic method is proposed to destroy, extract and replace the robust watermark successfully when the fragile watermarks cannot be detected.
(3) An improved watermarking scheme is proposed to resist the attack methods proposed in this paper.
(4) Further test the security and performance of the improved scheme.

The rest of this paper is organized as follows: Section 2 presents the specific content of the original dual watermark framework. Through analyzing the security vulnerabilities of the original framework, Section 3 provides the attack methods and experimental results against the original framework. Section 4 puts forward the improvement measures and gives the simulation results of the attack test. Section 5 summarizes the full text.

2. Contents of the Original Scheme Framework

2.1. Original Framework

The original scheme protects the copyright of the media image by embedding a robust watermark image in the transform domain using an inter-block coefficient difference algorithm and achieves the content authentication of the image by embedding a fragile watermark image in the spatial domain. In the case that the carrier image is a color image, the carrier image is first divided into three color space channel images of R, G and B, and then the robust watermark with double encryption is embedded in channel B, and the fragile watermark with double encryption using the same encryption technology is embedded in channel G, and the channel R remains unchanged. The original dual watermark frame is shown in Figure 1, where the carrier image is a 512 $\times$ 512 Lena color image, the robust watermark and fragile watermark are, respectively, a 64 $\times$ 64 binary panda image and character image.

2.2. Watermark Preparation

Two different encryption technologies are used to protect the security of embedded information, namely a robust watermark and a fragile watermark. First, the Arnold algorithm with the iteration number 'K_1' as the unique key is used for encryption, and then a new encryption technology is used to encrypt each row and column of the watermark encrypted by the Arnold algorithm by expanding a 32-bit key (K_2 = a1a2a3 ... a32) into a 64-bit encoding sequence. The watermark information W_e after two-stage encryption is embedded in the B-channel image. Algorithm 1 gives a two-level encryption algorithm for watermark information, and Figure 2 describes the specific process of encryption.

2.3. Embedding Algorithm

2.3.1. Embedding Robust Watermark

The carrier image of $M \times N$ is divided into the three channels of R, G and B. The robust watermark after double encryption is a binary image of P $\times$ Q. The first step is to

perform a single-stage Haar wavelet transform on the B-channel image. The reason why multi-level wavelet transform is not used is because of its complexity. Then, decompose the transform domain image into four sizes, all of which are $M/2 \times N/2$ sub-bands: LL, LH, HL and HH. The watermark data is embedded in the LL sub-band that contains the basic information of the image. The specific embedding process is given by Algorithm 2.

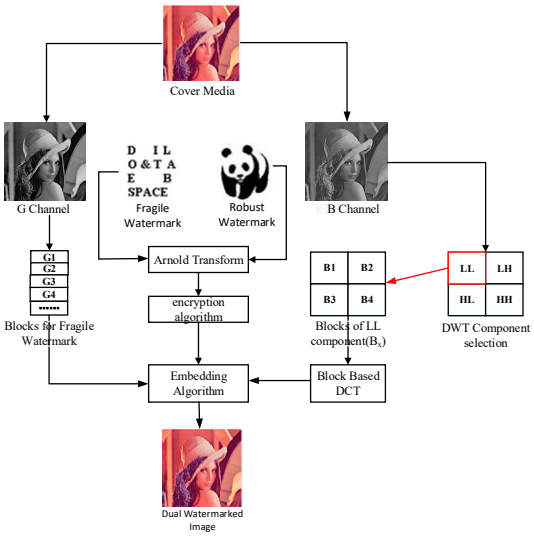


Figure 1. Framework of the original scheme.

Algorithm 1: Two-Level Encryption Algorithm

Input: Watermark image W, Key: iteration number K_1, $K_2 = a1a2a3 \cdots a32$
Output: Double-encrypted watermark image W_e
1: The watermark image W is encrypted by Arnold to obtain W_A, where the parameters a and b are known, (x',y') are the coordinates after the pixel (x, y) transformation, and N is the order of the pixel matrix:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{bmatrix} 1 & a \\ b & \mathrm{ab}+1 \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \bmod(N) \tag{1}$$

2: Convert the 32-bit key K_2 to Gray code and flip it to obtain the key $K_3 = b1b2b3 \ldots b32$ and divide it into $K3o = b1b3b5 \ldots b31$ and $K3e = b2b4b6 \ldots b32$ according to the parity bit.
3: Initialize the feedback bits Bo and Be, and perform the following operations to obtain Se and So:

$$\begin{aligned} Se &= b1 + b3 + Bo \\ So &= b2 + b4 + Be \end{aligned} \tag{2}$$

4: Perform the following XOR operation and concatenate the result to form an 8-bit sequence:

$$\begin{aligned} C11 &= \mathrm{a1} \oplus \mathrm{a2} \oplus Se \\ C12 &= \mathrm{a1} \oplus \mathrm{a2} \oplus So \\ C13 &= \mathrm{a3} \oplus \mathrm{a4} \oplus Se \\ C14 &= \mathrm{a3} \oplus \mathrm{a4} \oplus So \end{aligned} \tag{3}$$

5: Obtain the new values of Be and Bo and use them for the next cycle:

$$\begin{aligned} Be &= Se(1) \oplus Se(2) \\ Bo &= So(1) \oplus So(2) \end{aligned} \tag{4}$$

6: Repeat steps 3–5 seven times, and use $K3o$, $K3e$ and K_2 to form a 64-bit key encoding sequence K_4 to encrypt the watermark image W_A after the Arnold encryption:

$$\begin{aligned} W_{Ar} &= \mathrm{bitxor}(W_A(r), K_4) \\ W_{\mathrm{e}} &= \mathrm{bitxor}(W_{\mathrm{Ar}}(c), K_4) \end{aligned} \tag{5}$$

7:End procedure.

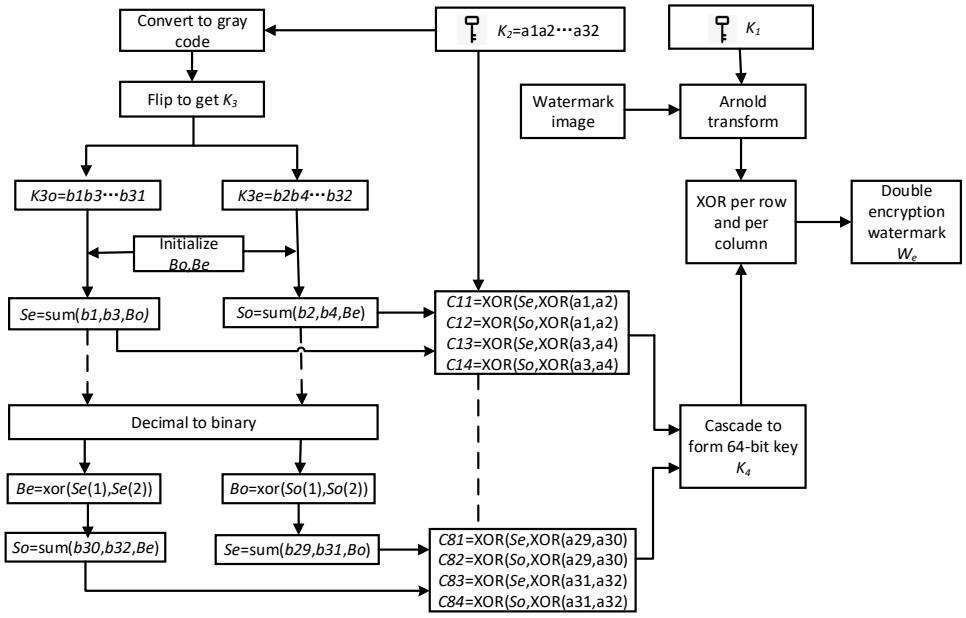


Figure 2. Watermark information encryption flow chart.

Algorithm 2: Robust Watermark Embedding Algorithm

Input: Double-encrypted robust watermark image W_{eb}, B-channel image
Output: B-channel image B_w embedded with robust watermark ciphertext
1: Subtract 128 from the pixel value of the B-channel image so that the pixel value range is in $[-128, 128]$. The single-stage wavelet transform is used to divide it into LL, LH, HL and HH sub-bands, and the size of each sub-band is $M/2 \times N/2$:

$$[\mathrm{LL}, \mathrm{LH}, \mathrm{HL}, \mathrm{HH}] = \mathrm{dwt2}(I, '\mathrm{haar}') \tag{6}$$

2: The LL sub-band is decomposed into n 8×8 non-overlapping blocks:

$$n = \frac{M/2 \times N/2}{8 \times 8} \tag{7}$$

Each 8×8 sub-block is further divided into four 4×4 sub-blocks. Therefore, the total number of bits embedded in the B-channel image is $4n$.
3: Each 4×4 block is transformed by DCT. Calculate the difference between the two pre-defined DCT coefficients selected from a pair of 4×4 adjacent blocks. Double-encrypted robust watermark sequence bits are embedded in DCT-transformed sub-blocks by modifying the difference. The difference (*Dif*) between a selected pair of DCT coefficients from sub blocks B_a and B_b is defined in Equation (8):

$$Dif = B_a(i, j) - B_b(l, m) \tag{8}$$

where $B_a(i,\ j)$ and $B_b(l,\ m)$ are the DCT coefficients chosen within a sub-block having different coordinates.
4: The difference is in any of the four predefined different areas. The occupied area is determined by the actual difference between the two comparison coefficients and the watermark bits to be embedded. In order to embed the '0' bit, it is set in area 2 or area 4. Similarly, for embedded bit '1', it is located in region 1 or region 3.
5: Perform IDCT transformation on the modified image block, and then perform IDWT of the modified approximation coefficient LL and the original detail coefficient (HL, LH and HH sub-bands).
6: Add 128 to the pixel value of the image after inverse transformation, so that the pixel value range is within [0, 255], that is, the B-channel image B_w embedded with robust watermark ciphertext is obtained.
7: End procedure.

2.3.2. Embedding Fragile Watermark

The fragile watermark W_{ec} after double encryption is embedded in the G-channel image. First, the channel image is divided into 8×8 non-overlapping blocks, then embed the watermark bit into the whole G_{xy} block by removing a predefined pixel point $G_{xy}(g, h)$. Then, the pixel $G_{xy}(g, h)$ is used to embed the duplicate version of the same watermark bit. The same watermark is embedded twice in such a block in order to achieve the tamper location of potential attackers through subsequent extraction operations. The specific embedding process is given by Algorithm 3.

Algorithm 3: Fragile Watermark Embedding Algorithm

Input: Double-encrypted fragile watermark image W_{ec}, G-channel image
Output: G-channel image G_w embedded with fragile watermark ciphertext
1: The G-channel image is divided into non-overlapping block of 8×8 and recorded as G_{xy} (x, y = 1, 2, 3 ... $(M \times N)/(64 \times 64)$), the block mean value after removing one predefined pixel is calculated:

$$m = \frac{sum(G_{xy}) - G_{xy}(g,h)}{63} \tag{9}$$

2: Adjust the block mean value according to the embedded watermark bit:

$$\mathrm{mod}(m/\delta, 2) = \begin{cases} 1; for & w = 1 \\ 0; for & w = 0 \end{cases} \tag{10}$$

3: When the formula condition is not satisfied, iterate and add 1 to the mean value m until it is satisfied. Otherwise, the modified mean value will be recorded as m_n:

$$\Delta = m_n - m \tag{11}$$

4: Calculate the duplicate version G'_{xy} and pixel of the watermark block $G_{xy}(g,h)$. The watermark bit is embedded in the LSB of the pixel value:

$$\begin{aligned} G'_{xy} &= G_{xy} + \Delta \\ G'_{xy}(g,h) &= G_{xy}(g,h) - \Delta \end{aligned} \tag{12}$$

5: End procedure.

2.4. Watermark Extraction

Watermark extraction is the inverse process of watermark embedding. Since the original scheme framework is blind, it does not need the participation of the original watermark. After converting the double watermark image into an RGB color space model and dividing it into three channels, select the B channel to extract the double encrypted robust watermark and decrypt it, select the G channel to extract the fragile watermark ciphertext twice and compare them. If the extracted watermarks are identical, it means that the image block has not been tampered with; otherwise, it means that the image block has been attacked. In this way, the possible tampering attacks of potential attackers can be located.

3. Security Analysis and Attack on the Original Scheme Framework

3.1. Security Analysis

The original scheme uses the RGB color space model to embed the robust watermark in the transform domain of the B channel through DWT and DCT transformation, and the fragile watermark is embedded twice in the space domain of the G channel. The R channel is intact. In addition, in order to ensure the security of embedded watermark information, the Arnold algorithm and a new key information expansion technology are used to encrypt the watermark. The robust watermark is characterized by robustness, which can resist a certain degree of signal processing without destroying the watermark itself; the fragile

watermark is sensitive and can be used to verify the integrity of the protected content and tamper detection and location.

Through the security analysis of the original watermarking framework, it can be found that although the embedding method of robust watermarking combines DWT and DCT transformation, and after the carrier image is single-stage DWT transformed, the LL sub-band is embedded into the block with rich details adaptively by block DCT and the quantization step is also added. However, in essence, it is still based on the basic idea of jitter modulation, that is, the quantization interval is modulated according to the watermark bit. However, as early as 2006, the method of jitter modulation was cracked by the scheme [36] proposed by Lu et al. using forgery attack. However, this paper gives some attacks from another perspective. In addition, there are two obvious security vulnerabilities in the original scheme framework, as shown below.

3.1.1. Security Vulnerability 1

In the original framework, the watermarking information is protected by the dual encryption method of scrambling before encrypting, but it cannot prevent the robust watermarking from being destroyed. The Arnold transform takes the number of iterations as the unique key, and the new encryption method is not significant in expanding the 32-bit key ($K_2 = a_1a_2a_3 \cdots a_{32}$) to a 64-bit key ($K_4 = c_1c_2c_3 \cdots c_{64}$) and uses it to encrypt each row and column of the watermark image. Attackers can directly treat it as a 64-bit one-dimensional key ($K_4 = c_1c_2c_3 \cdots c_{64}$) to XOR the watermark after Arnold scrambling according to the following formula:

$$W_A(i,j) \oplus K_4(i) \oplus K_4(j) = W_e(i,j) \tag{13}$$

Here, i, j is the coordinate of the watermark pixel and $i, j = 1, 2, 3 \cdots 64$. In the second encryption stage, assuming that the attacker has two robust watermarks and images embedded with watermarks, the attacker can use the known plaintext attack to obtain the XOR result of the scrambled images W_{A1} and W_{A2} by XOR the two robust watermarked images W_1, W_2 and the extracted ciphertext image W_{e1} and W_{e2} by removing the equivalent XOR key. If there is a scrambling relationship between the XOR result of two plaintext watermarks and the XOR result of two scrambled images, it is equivalent to obtaining a plaintext watermark and the scrambled image. Attackers can predict the number of iterations to obtain the scrambling key, further obtain the equivalent XOR key, achieve the extraction and replacement of robust watermark, and tamper with copyright information.

3.1.2. Security Vulnerability 2

Although the original scheme claims to be able to achieve tamper localization, we find that the tamper localization claimed by it is only effective for the host image embedded with double watermarks. More seriously, the removal attack shown in [35] is also not effective for the double watermarks image. The reason for this phenomenon is that there are flaws in the tamper localization method for fragile watermarks. For the removal attack, the embedded watermark contained in the removed partial channel image can no longer be extracted, and the removal will not have any impact on the distributed embedded fragile watermark ciphertext and its copies in the remaining part of the image so that the fragile watermarks 1 and 2 extracted successively are identical, and the tampering localization is invalid. Secondly, the RGB color space model is composed of the three most sensitive color lights, red, green and blue, which are superimposed in different proportions by the additive mixing method. The three channels are related to each other and have spatial independence. When one channel is attacked, the other two channels will not be affected by the attack (Figure 3).

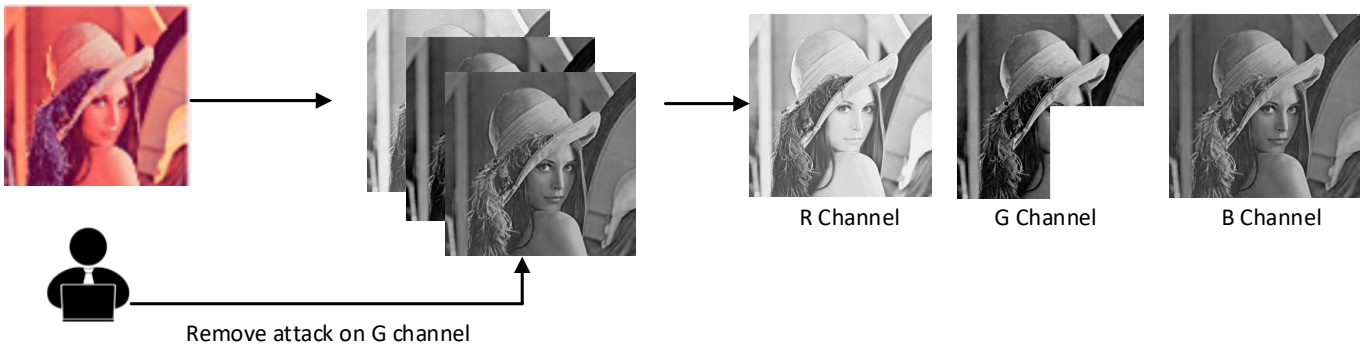


Figure 3. Schematic diagram of single-channel attack.

Many scholars embed robust watermarks and fragile watermarks in different channels to achieve copyright protection and content authentication of carrier information at the same time, so if each channel is independent of the other, it is possible that one channel will be attacked separately. For the robust watermark and fragile watermark embedded in the B channel and the G channel, respectively, there is no correlation between them, so when the attacker attacks the B channel embedded in the robust watermark alone, it will not affect the fragile watermark embedded in the G channel, and the fragile watermark cannot detect the attacker's tampering with the robust watermark. Then, the attacker can extract and replace the robust watermark of channel B while avoiding the fragile watermark detection of channel G.

3.2. Attack

As we all know, based on the Kerckhoff principle [37], the security of a cryptographic system depends entirely on the key, and not the complexity of the cryptographic system. In other words, the attacker knows everything about the original password scheme and has access to the encryption mechanism except the key. The goal of cryptographic analysis is to obtain all or part of the key or the equivalent key. After obtaining the key, the attacker can perform the following: unauthorized decoding, unauthorized extraction and unauthorized embedding. After understanding the algorithm and framework of the original encryption system, anyone who wants to attack the encryption system can easily analyze the potential security vulnerabilities in the original encryption system, thus disclosing the key information in the original encryption system. Based on the above security vulnerabilities of the original dual watermark scheme, we propose the following attack methods. Figure 4 is our proposed framework for attacking the original dual watermark scheme.

3.2.1. Destroy Robust Watermark without Being Detected by Fragile Watermark

1. The purpose of the attacker is to destroy the robustness of the dual watermark scheme by destroying the robust watermark. Because the embedding algorithm of the robust watermark in the original scheme is essentially based on the main idea of jitter modulation, and according to the correlation between block coefficients, after DWT transformation of the B-channel image of $M \times N$, LL sub-band of $M/2 \times N/2$ is divided into 8×8 sub-blocks, and then each sub-block is divided into 4×4 and conduct DCT transform, and the ciphertext robust watermark is embedded by modifying the DCT coefficients of some selected B channels. Therefore, after knowing the specific details of the watermark embedding algorithm, the attacker can destroy the robust watermark by modifying some DCT coefficients again.
2. DCT transform has good decorrelation, usually with 8×8-pixel blocks. The smaller the block unit is, the lower the complexity of the image algorithm is. However, the performance of DCT decorrelation is weakened, which can easily cause obvious image block effect. The figure below shows the coefficient distribution diagram

after a 4 × 4-pixel block is transformed by DCT which is from a size of $M/2 \times N/2$ LL sub-band. One of the DCT coefficient blocks is recorded as $B_{m,n}(x, y)$ ($1 <= m, n <= M/8, 1 <= x, y <= 4$), and the pixel points in the upper left corner are $B_{m,n}(1,1)$ are DC coefficients, that is, the low-frequency signal part of this pixel block, which concentrates the main energy of the original image. The rest is AC coefficient, that is, a high-frequency signal part. The closer it is to the lower right corner, the smaller its value is and close to 0 (Figure 5).

3. The original scheme selects two adjacent 4 × 4 small pieces from an 8 × 8 sub-block and predefines the difference between the two DCT coefficients in the adjacent blocks. Then, the difference *Dif* is calculated according to the formula 15. Figure 6 is the visualization of the 4 × 4 sub-block difference direction matrix:

$$Dif = \begin{cases} LR : B_{m,n}(x,y) - B_{m,n+1}(x,y); for \bmod(m,2) = 1 \&\& \bmod(n,2) = 1 \\ UD : B_{m,n}(x,y) - B_{m+1,n}(x,y); for \bmod(m,2) = 1 \&\& \bmod(n,2) = 0 \\ RL : B_{m,n}(x,y) - B_{m,n-1}(x,y); for \bmod(m,2) = 0 \&\& \bmod(n,2) = 0 \\ DU : B_{m,n}(x,y) - B_{m-1,n}(x,y); for \bmod(m,2) = 0 \&\& \bmod(n,2) = 1 \end{cases} \quad (14)$$

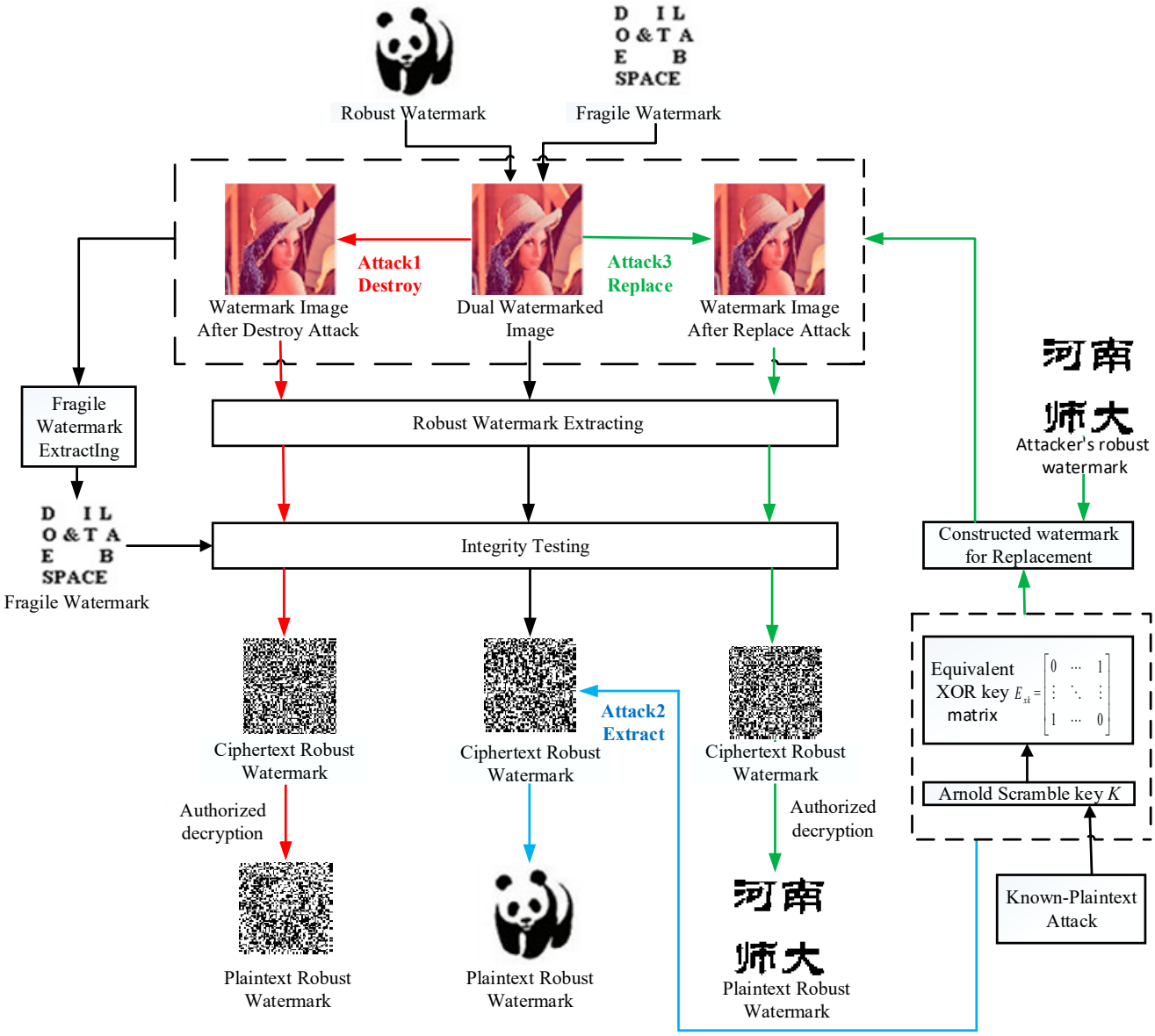


Figure 4. The attack framework of the original dual watermark scheme.

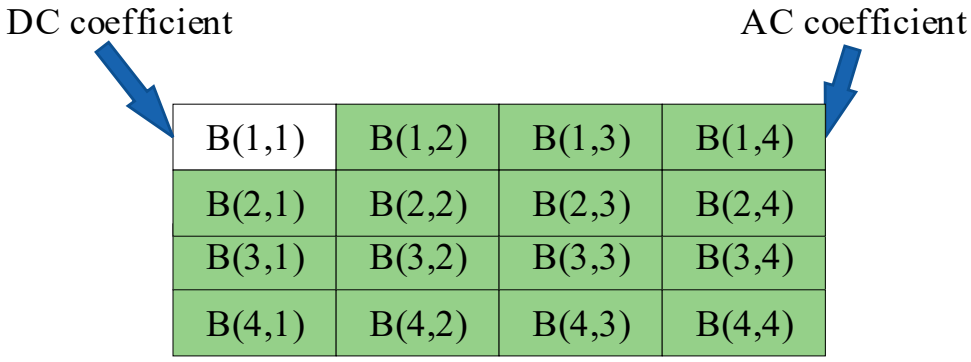


Figure 5. 4 × 4-pixel block DCT coefficient distribution.

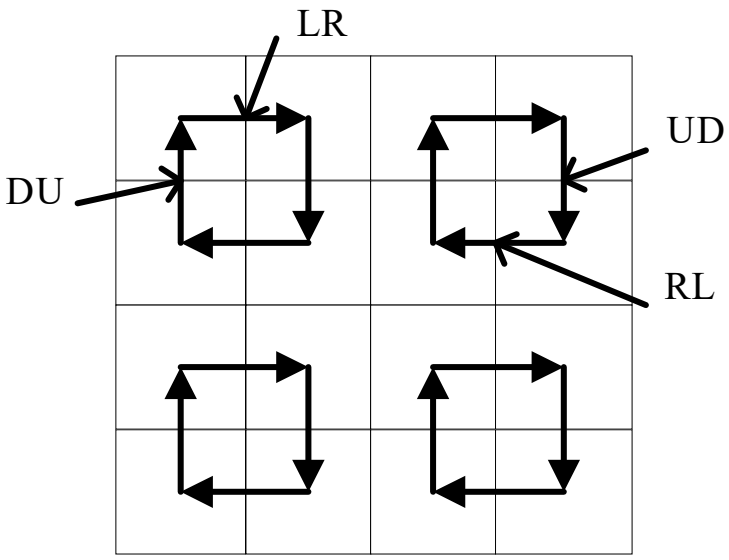


Figure 6. Difference direction matrix of 4 × 4 sub-blocks.

The original scheme completes the embedding of the ciphertext robust watermark by modifying the difference *Dif* so that it is located in four predefined different regions. The four different regions are determined by the threshold parameter T, the embedding factor E and the actual difference, and the parameter u is the iterative correction amount. According to the Kerckhoff principle [37], the attacker knows all the details of the watermark system. Then, the attacker can modify the coefficients of the embedded watermark bit again to destroy the extracted robust watermark. As shown in Figure 7, the attacker can achieve the purpose of destroying the robust watermark by modifying the embedded watermark bit '1' or '0' to a non-original embedded region, even if there is an isolation zone between the four embedded regions. When the embedded watermark bit is '0', the difference *Dif* is modified out of the original regions 2 and 4. When the embedded watermark bit is '1', the difference *Dif* is modified out of the original regions 1 and 3. Algorithm 4 describes the specific process of destroying robust watermarks in detail, and Figure 8 shows the simulation results of destroying robust watermarks.

It can be seen from the image (f) in the simulation experiment result (Figure 8) of destroying the robust watermark that the decrypted robust watermark (c) after the attack is significantly different from the original embedded and decrypted robust watermark (e), which means that the attack is effective, and the original robust watermark cannot be extracted from the carrier image after the damage attack. However, at this time, the extracted and decrypted fragile watermark (d) is exactly the same as the original embedded watermark, indicating that it has not detected the damage attack to the robust watermark.

Algorithm 4: Destroy Robust Watermark

Input: double watermark image
Output: extracted robust watermark and fragile watermark
1: Convert the double watermark image into RGB color space image and split it into R-, G- and B-channel images.
2: Subtract the pixel value of the B-channel image by 128 and perform DWT transformation, then divide the LL sub-band into 8×8 small pieces, and then divide each 8 × 8 small pieces into 4 × 4, and finally perform DCT transformation.
3: Calculate the difference *Dif*, and change the DCT coefficient $B_{m,n}(x, y)$ selected in the embedding algorithm. When the iterative correction u keeps the embedded watermark bit unchanged, the difference *Dif* is modified out of the original region.

IF mark(i) == 1 **then**
IF $Dif > T + E$
while $Dif > T + E$
$B_{m,n}(x, y) = B_{m,n}(x, y) - u$
$Dif = Dif - u$
End while
Elseif $(Dif > -T/2)$ && $(Dif < -E)$
While $Dif < -E$
$B_{m,n}(x, y) = B_{m,n}(x, y) + u$
$Dif = Dif + u$
End while
Elseif $(Dif < -T/2)$ && $(Dif > -T + E)$
While $Dif > -T + E$
$B_{m,n}(x, y) = B_{m,n}(x, y) - u$
$Dif = Dif - u$
End while
Else
$B_{m,n}(x, y) = B_{m,n}(x, y)$
$Dif = Dif$
End
Else mark(i) == 0 **then**
If $(Dif > T/2)$ && $(Dif < T - E)$
While $dif < T - E$
$B_{m,n}(x, y) = B_{m,n}(x, y) + u$
$Dif = Dif + u$
End while
Elseif $(Dif < T/2)$ && $(Dif > E)$
While $Dif > E$
$B_{m,n}(x, y) = B_{m,n}(x, y) - u$
$Dif = Dif - u$
End while
Elseif $Dif < -T - E$
While $Dif < -T - E$
$B_{m,n}(x, y) = B_{m,n}(x, y) + u$
$Dif = Dif + u$
End while
Else
$B_{m,n}(x, y) = B_{m,n}(x, y)$
$Dif = Dif$
End
End IF

4: Apply inverse DCT to each block and combine the original HL, LL and LH sub-bands for IDWT operation.
5: Add 128 to the pixel value of the attacked B-channel image and combine it with the G and B channels.
6: Split the combined image, extract the robust watermark from the B-channel image to check the robustness, and extract the fragile watermark from the G-channel image to check the tamper location.
7: End procedure.

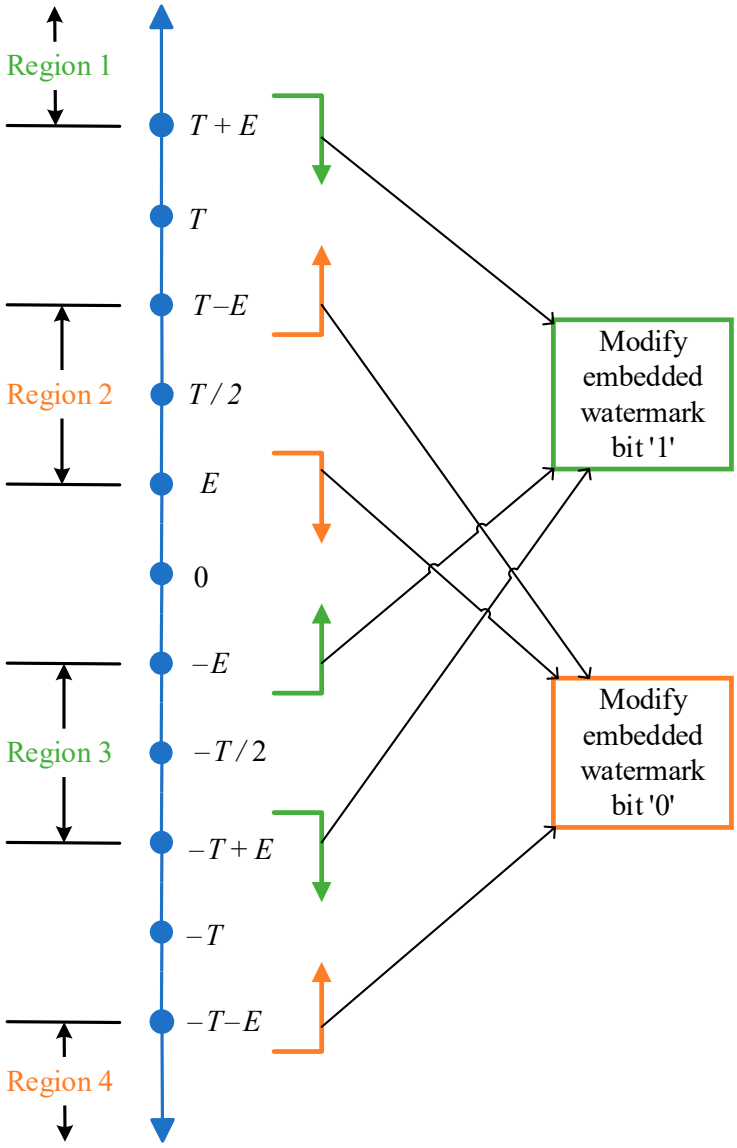


Figure 7. Modification of coefficient for embedding a watermark bit '1' or '0'.

(**a**) (**b**) (**c**) (**d**) (**e**) (**f**)

Figure 8. Simulation experiment of destroying robust watermark: (**a**) carrier image after being attacked; (**b**) the robust watermark ciphertext extracted from (**a**); (**c**) decrypted robust watermark; (**d**) the fragile watermark extracted and decrypted from (**a**) (1 is the same as 2); (**e**) the original decrypted robust watermark; (**f**) XOR image representing (**c**,**e**).

3.2.2. Extracting Robust Watermark without Being Detected by Fragile Watermark

Attackers usually steal copyright information by extracting a robust watermark. The security of the original dual watermark scheme mainly depends on the keys K_1 and K_4. Before embedding, the robust watermark in the original scheme undergoes the double encryption operation of Arnold scrambling and extended key ($K_4 = c_1c_2c_3 \cdots c_{64}$) XOR processing. However, this does not provide complete security protection. An attacker can obtain the key through known plaintext attacks and then steal copyright information. The attacker can use two known two watermarked images and their plaintext robust watermark W_1 and W_2, extract two ciphertext images W_{e1} and W_{e2}, respectively, from the watermarked images, and XOR operations are performed on the plaintext robust watermarks W_1 and W_2, and the ciphertext robust watermark W_{e1} and W_{e2}, respectively. The specific process of cryptanalysis is shown in Figure 9.

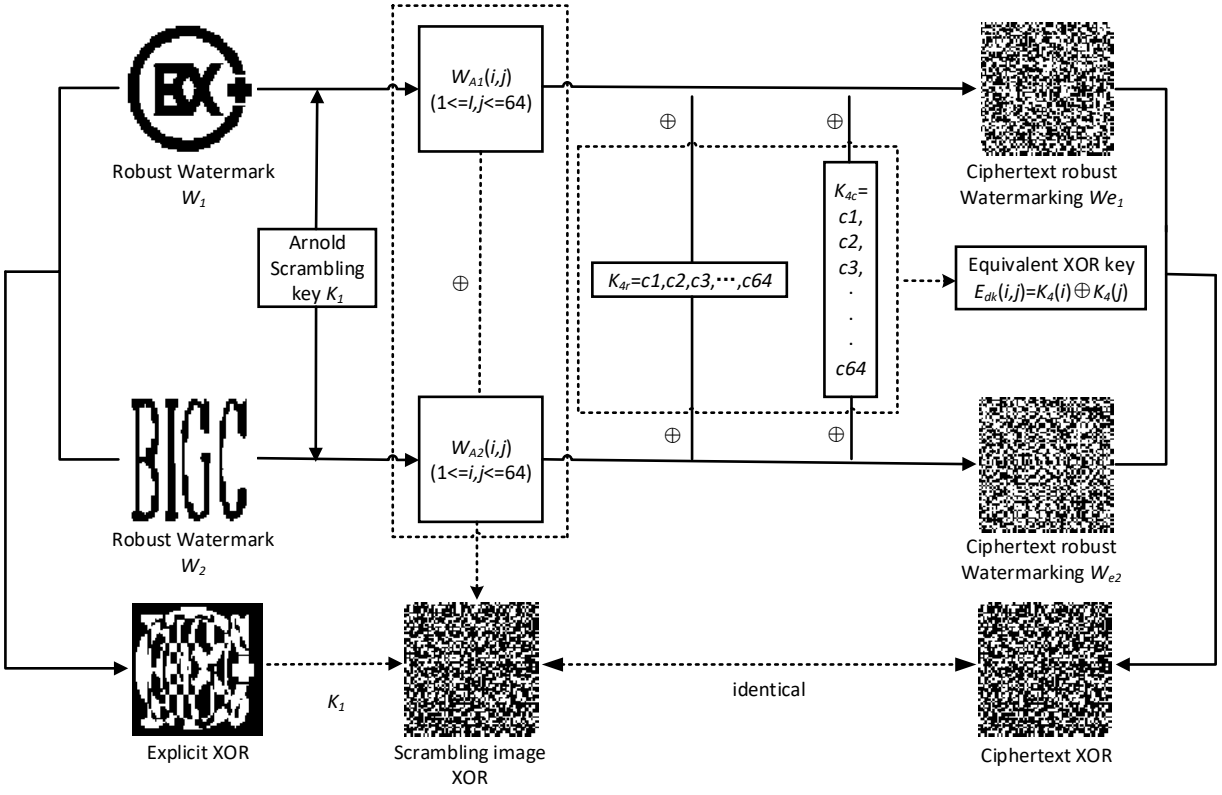


Figure 9. Watermark double encryption process cipher analysis.

In the secondary encryption stage, the scrambled images W_{A1} and W_{A2} are unknown, and the equivalent XOR key $E_{xk}(i,j)$ is known, expressed as $E_{xk}(i,j) = K_4(i) \oplus K_4(j), (1 <= i,j <= 64)$, using the formula 15:

$$\begin{aligned} W_{A1}(i,j) \oplus K_4(i) \oplus K_4(j) &= W_{e1}(i,j) \\ W_{A2}(i,j) \oplus K_4(i) \oplus K_4(j) &= W_{e2}(i,j) \end{aligned} \tag{15}$$

By eliminating the equivalent XOR key $E_{xk}(i,j)$, we can know that the XOR result of scrambling images W_{A1} and W_{A2}, is the same as the XOR result of ciphertext images W_{e1} and W_{e2}.

$$W_{A1}(i,j) \oplus W_{A2}(i,j) = W_{e1}(i,j) \oplus W_{e2}(i,j) \tag{16}$$

In the scrambling phase, the attacker is equivalent to obtaining a plaintext robust watermark image and its scrambling result image. By predicting the number of iterations, K_1 in the scrambling phase can be obtained. In the XOR stage, the attacker can obtain the equivalent XOR key = $E_{xk}(i,j)$, according to one of the known robust watermarks W_1 and

W_2 and its ciphertext image. Attackers can extract the robust watermark W_b by decrypting the extracted encrypted robust watermark W_{eb} after obtaining K_1 and $E_{xk}(i,j)$, and realize the theft of the robust watermark. Algorithm 5 describes the extraction process of the robust watermark in detail, and the simulation results are shown in Figure 10.

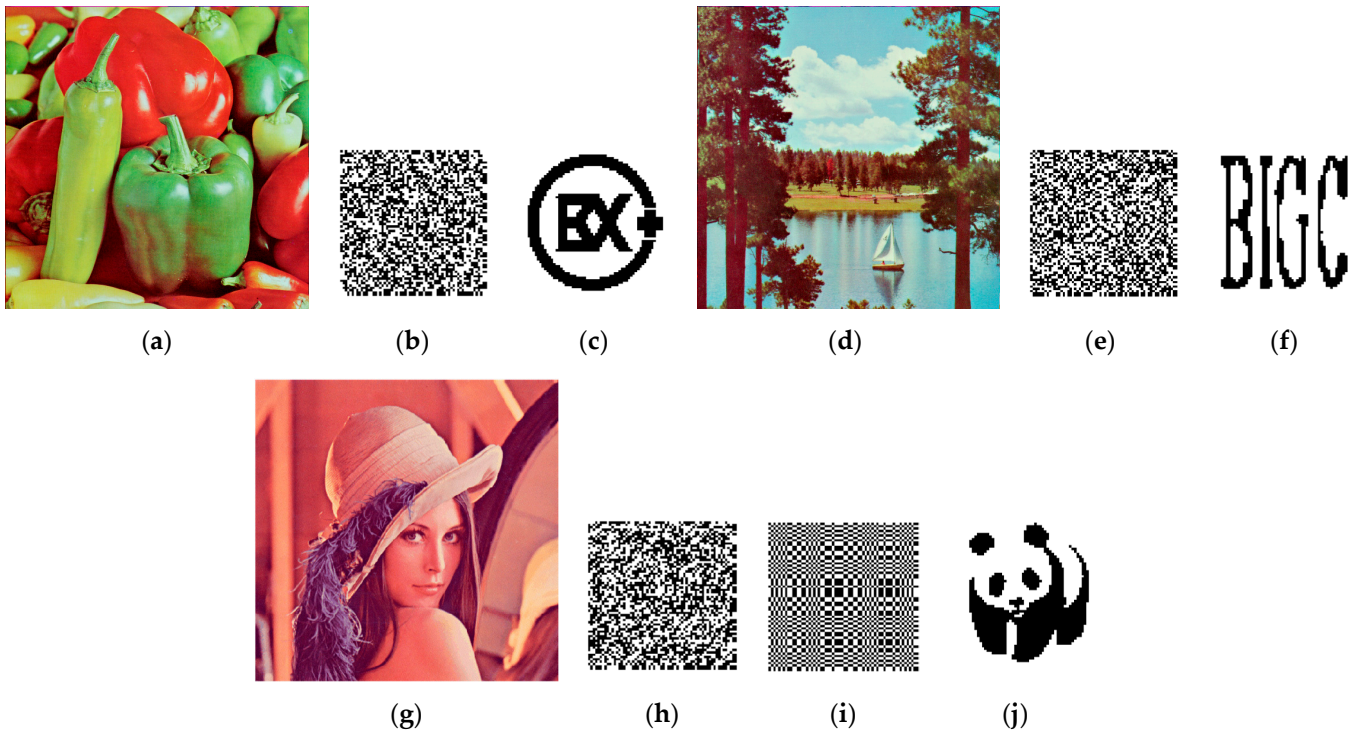


Figure 10. Simulation results of extracting robust watermark (**a**) and (**d**) are the watermarked images known to the attacker; (**b**) the ciphertext robust watermark extracted from; (**c**) the plaintext robust watermark of (**b**); (**e**) the ciphertext robust watermark extracted from (**d**); (**f**) the plaintext robust watermark extracted from (**e**); (**g**) the double watermark image to be attacked; (**h**) the extracted ciphertext robust watermark; (**i**) the equivalent XOR key matrix image obtained from the known plaintext attack; (**j**) the plaintext robust watermark obtained by the attacker.

Algorithm 5: Extracting Robust Watermark W_b

Input: double watermark image, the watermark image known by the attacker and the robust watermark W_1 and W_2 contained therein
Output: original embedded robust watermark W_b
1: Convert the double watermark image into an RGB color space image and split it into R-, G- and B-channel image, and extract the double-encrypted robust watermark W_{eb} from the B channel.
2: The attacker extracts the encrypted robust watermarks W_{e1} and W_{e2} from the known watermarked image, and according to the robust watermarks W_1 and W_2, obtains formula 15 from Figure 9, and then obtains the scrambling key K_1.
3: The attacker select one of the known robust watermarks W_1 and W_2 to scramble it with using the scrambling key K_1, and obtains the corresponding scrambling image W_{A1} or W_{A2}, and then performs the following operations to obtain the equivalent XOR key:

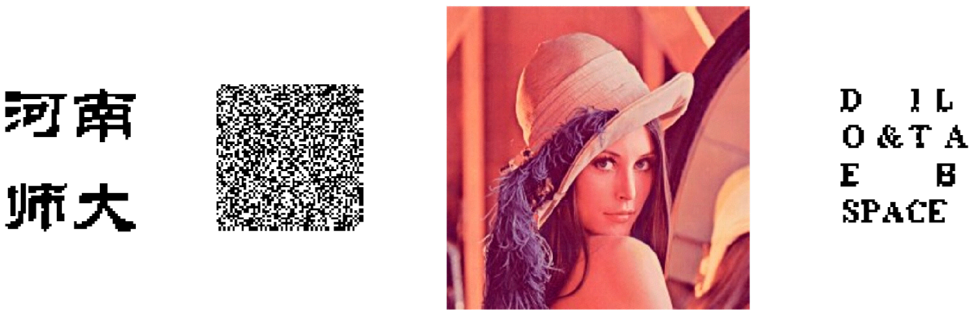
$$\begin{aligned} E_{xk}(i,j) &= W_{A1}(i,j) \oplus W_{e1}(i,j) \\ &= W_{A2}(i,j) \oplus W_{e2}(i,j) \end{aligned} \tag{17}$$

4: The attacker uses the scrambling key K_1 and the equivalent XOR key obtained above to decrypt the double-encrypted robust watermark W_{eb} and extract the original embedded robust watermark W_b.
5: End procedure.

In the simulation experiment result (Figure 10) of extracting robust watermark, the attacker uses the known plaintext attack, according to the known watermarked images (a), (d) and the copyright information contained therein, that is, the robust watermark W_1 and W_2 (images (c) and (f)), and extracts the ciphertext robust watermark W_{e1} and W_{e2} from the known watermarked images. The Arnold scrambling relation is explicitly defined by XOR two plaintext robust watermark images W_1 and W_2 and two ciphertext robust watermark images W_{e1} and W_{e2}, respectively. The key K_1 is predicted, and the equivalent XOR key (image (i)) in the XOR encryption phase can be obtained by using one group of plaintext robust watermark images. At this time, the watermark encryption method of the original scheme has been completely cracked. The robust watermark image (j), extracted by the attacker using the cracking key, is identical to the original embedded robust watermark, indicating that the attacker can successfully extract the robust watermark.

3.2.3. Replacing Robust Watermark without Being Detected by Fragile Watermark

If the attacker can steal the robust watermark, then the robust watermark will also face a great risk of being tampered with. After obtaining the scrambling key K_1 and the equivalent XOR key $E_{xk}(i, j)$ for double encryption of the robust watermark, the attacker can arbitrarily make a watermark image containing his own information and then generate a ciphertext watermark image to replace the original encrypted robust watermark embedded in the carrier image. In other words, attackers can arbitrarily replace the original robust watermark to tamper with copyright information. Algorithm 6 describes the replacement process of the robust watermark in detail, and the simulation results are shown in Figure 11.



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

Figure 11. Simulation results of replacing robust watermark (**a**) the watermark image owned by the attacker; (**b**) Self-produced ciphertext watermark image by the attacker; (**c**) the double watermark image with tampered copyright information; (**d**) the fragile watermark image extracted from (**c**).

Algorithm 6: Replacing Robust Watermark W_b

Input: double watermark image, attacker's own watermark image W_s
Output: double watermark image of modified copyright information
1: Convert the double watermark image into an RGB color space image and split it into R-, G- and B-channel images and extract the double-encrypted robust watermark *Web* from the B channel.
2: Use the scrambling key K_1 obtained in the previous section to perform Arnold scrambling on W_s to obtain W_{AS}, and the equivalence XOR key is re-encrypted to produce a robust watermark ciphertext image W_{es}:

$$W_{es}(i, j) = W_{As}(i, j) \oplus E_{xk}(i, j) \tag{18}$$

3: Embed the tampered robust watermark ciphertext image W_{es} into channel B.
4: The G and R channels remain unchanged, and the B channel embedded with the replacement watermark is combined to obtain the double watermark image with the tampered copyright information.
5: End procedure.

The attacker further uses the scrambling key K_1 and the equivalent diffusion key $E_{xk}(i, j)$ to produce a ciphertext watermark image (b) with its own copyright information (a) in Figure 11 and generates a double watermark image (c) with the tampered copyright information. During this period, the fragile watermark (d) is not affected by the replacement attack. The above extraction and replacement of the robust watermark in channel B is not related to the fragile watermark in channel G, so the tamper detection and location of the fragile watermark can be avoided.

4. Improved Design Scheme with Resilience

4.1. Improvement Methods

The security flaws of the original double watermarking framework are firstly, the double encryption method of robust watermarking is not safe enough and can be cracked by attackers using known plaintext attacks. Secondly, the B channel embedded with the robust watermark and the G channel embedded with the fragile watermark are independent of each other. When the attacker attacks the B channel alone, the fragile watermark signal cannot be synchronized with the attacked robust watermark signal, and the attacker's destruction, extraction and replacement of the robust watermark cannot be detected. In order to overcome the above shortcomings, this paper, starting from improving the security of the framework and the performance of tamper location, has conducted in-depth research on improving the encryption security and tamper location effect and has made improvements in the following two aspects.

4.1.1. Encryption Method of Robust Watermark

In view of the fact that the Arnold scrambling algorithm used in the original scheme only uses the iteration number K_1 as the unique key to ensure security, with high risk, it is easy for attackers to decipher information after predicting the iteration number and cannot resist extraction and replacement attacks. Therefore, we adopt a new pixel spatial position scrambling algorithm. A chaotic system is highly sensitive to initial conditions and parameters. The logical map is a one-dimensional discrete chaotic system, which can be regarded as a function. The generated result is a one-dimensional aperiodic chaotic sequence, which is defined as follows:

$$X_n = F(x_{n-1}) = uX_{n-1}(1 - X_{n-1}) \tag{19}$$

in which, the value range of the control parameter u is {0, 4}, and the initial value X_0 is between 0 and 1. The value of X_n also ranges from 0 to 1. We take the initial value R_0 and the control parameter u as the key and generate the random number sequences $\{R_i\}$ of $M \times N$ different random numbers firstly, and sort it to obtain an ascending sequence $\{R'_i\}$, and then a coordinate sequence $\{P_i\}$ is predefined to record the positions of random numbers of $\{R'_i\}$ in the sequence $\{R_i\}$ in turn, and then the robust watermark is processed according to the coordinate sequence; thus, the scrambling robust watermark image W_p is obtained. In order to better improve the security of the encryption algorithm, we further conduct the diffusion operation on W_p. We set the value less than 0.5 in $\{R_i\}$ to 0 and the value in the range of $[0.5, 1)$ to 1 to generate a binary random sequence $\{D_i\}$ with a size of 64×64. The set initial value for the ciphertext robust watermark is $W_e(0) = D(0) \oplus W_p(0)$. The algorithm and its inverse algorithm when i from 1 to MN are as follows:

$$\begin{cases} W_{ei} = W_{e(i-1)} \oplus D_i \oplus W_{pi} \\ W_{pi} = W_{e(i-1)} \oplus D_i \oplus W_{ei} \end{cases} \tag{20}$$

Under the encryption method based on the above scrambling-diffusion algorithm, the attacker cannot crack the key, and it is effective to prevent the robust watermark from being extracted and replaced.

4.1.2. Establish an Association System between Robust Watermarking for Copyright Protection and Fragile Watermarking for Content Authentication

The fragile watermarking algorithm in the original scheme cannot effectively achieve tamper detection and location under a single-channel attack. The mutual independence of the three channels of RGB results means that the fragile watermark embedded in the G channel cannot synchronize the possible attacks on the robust watermark in the B channel. Therefore, an association system between the two watermarks is established to enable the fragile watermark to be subject to the same attacks as the robust watermark in the ciphertext in the B channel so as to better achieve accurate tamper localization against the robust watermark. This paper improves the fragile watermark by XOR operation of the fragile watermark and the ciphertext robust watermark to establish the relationship between them and improves the fragile watermark algorithm based on the original scheme. The improved fragile watermark and the original fragile watermark are successively embedded in the entire G-channel image block, and the original fragile watermark is embedded in the LSB of the copied version pixel $G'_{xy}(g, h)$. The tamper recovery of the robust watermark can also be realized under the attack method proposed in this paper. Figure 12 shows the design framework of our proposed improvement measures. Algorithm 7 describes the specific process. The simulation results of tamper recovery will be shown in the next section.

Algorithm 7: Improved Double Watermark Algorithm

Input: carrier image I, robust watermark, fragile watermark,
Output: improved double watermark image
1: Convert the carrier image into an RGB space image and split it into three channel images.
2: R←Logical map (R_0,u) //Generate random sequences R using chaotic system.
3: $R' \leftarrow$ sort(R) //Arrange the random number sequence R in ascending order.
4: $P \leftarrow R$, R' //Obtain the sequence P of the positions of the elements of R' in the sequence R.
5: $W_p \leftarrow$ P(W_b) //Obtain scrambling robust watermark W_p.
6: $W_{eb}(i) \leftarrow$ bitxor($W_e(i-1)$, bitxor(D_i,W_{pi})) //The binary sequence D is obtained to diffuse the scrambled robust watermark image W_p, and the ciphertext robust watermark W_{eb} is obtained.
7: $W_{mc} \leftarrow$ bitxor(W_c,W_{eb}) //Obtain the improved fragile watermark.
8: Embed the ciphertext robust watermark into channel B. The improved fragile watermark is embedded in the G channel, and the original fragile watermark is embedded in the LSB of the copied version pixel.
9: The three-channel image is combined and converted into an RGB image, which is the improved double-watermarked image.
10: End procedure

4.2. Simulation Test and Robust Watermark Recovery Process for the Proposed Attack Method

Generally speaking, in order to make the watermark scheme have the ability of recovery, it is necessary to embed some additional recovery information in the original carrier image. However, the more information is embedded, the more serious the distortion of the protected image is, and the image fidelity cannot be maintained [38]. The improved scheme, on the basis of improving the encryption method of the robust watermark, synchronizes the fragile watermark with the attack changes of the ciphertext robust watermark in channel B by establishing the association between the fragile watermark and the robust watermark. After the tampering is located, the tampered ciphertext robust watermark is recovered by using the fragile watermark extracted twice in the fragile watermark algorithm, and then the robust watermark is decrypted and recovered. It is not necessary to embed too much recovery information to meet the requirements of fidelity. We used the attack method proposed in the previous chapter to test the improved watermark framework. Algorithm 8 is the simulation test and the detailed description of the robust watermark recovery process. Figure 13 is the simulation results.

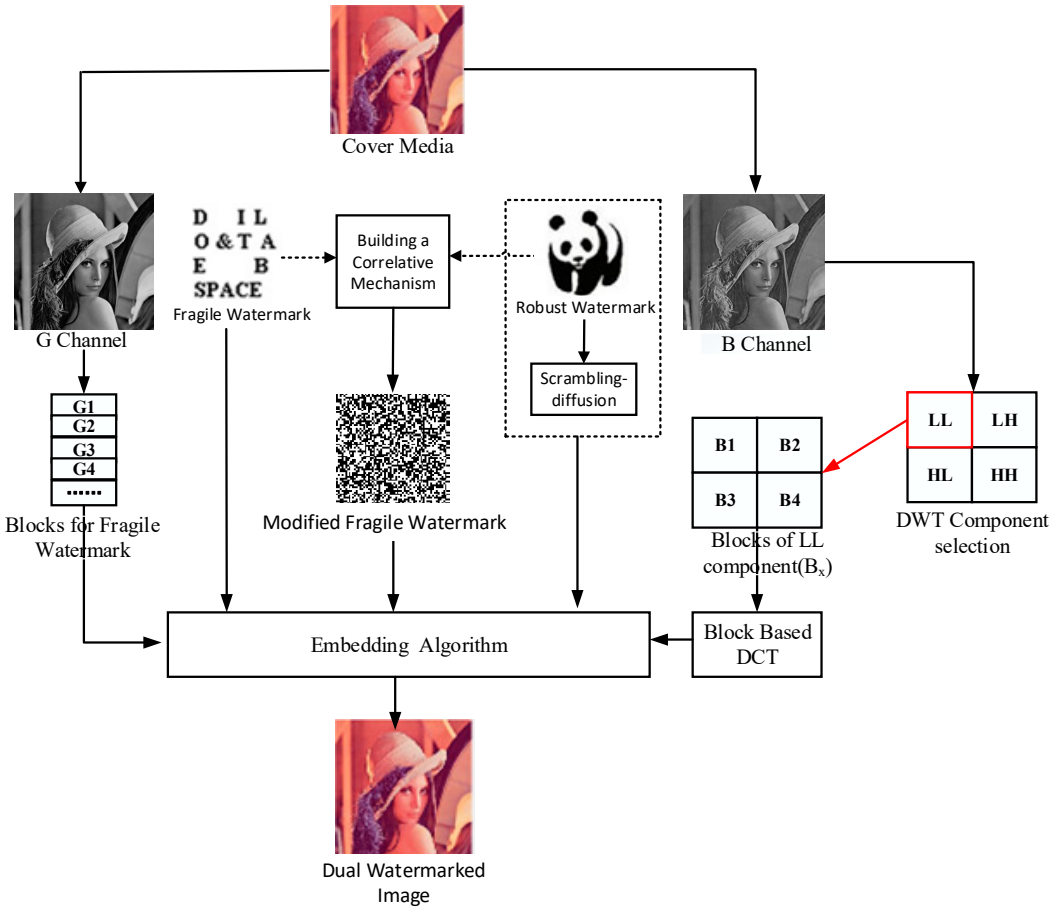


Figure 12. Improvement framework.

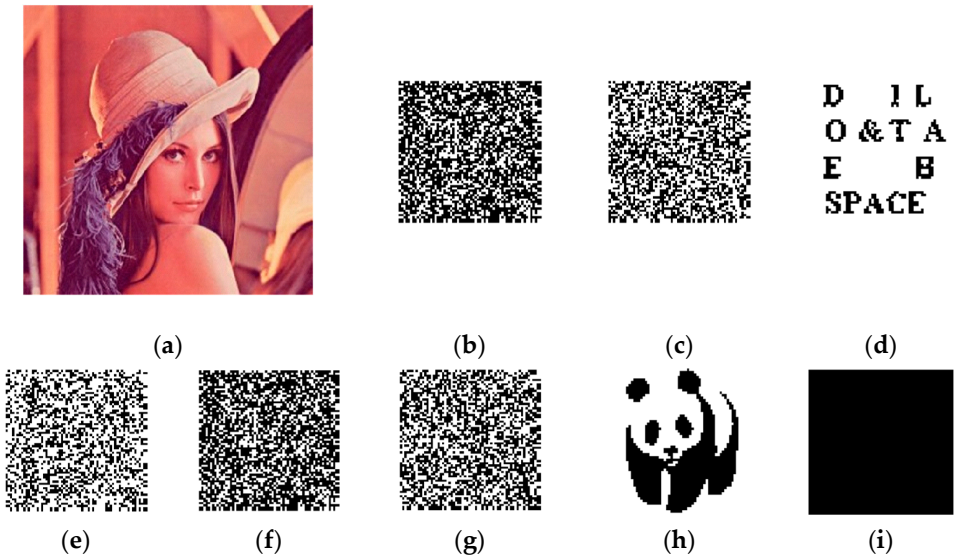


Figure 13. The simulation results of the improved scheme and the attack on the improved scheme (**a**) the combined image after the destroy attack; (**b**) the ciphertext robust watermark extracted from (**a**); (**c**) the improved fragile watermark extracted; (**d**) the fragile watermark extracted twice; (**e**) the XOR result of (**b**,**c**) , the attack signal of (**b**) is synchronized; (**f**) the location result of tampering; (**g**) the recovered ciphertext robust watermark; (**h**) the decrypted robust watermark of (**g**); (**i**) the XOR image of (**h**) and the original robust watermark, all black means that the recovered robust watermark is identical to the original watermark.

Algorithm 8: Simulation Test and Robust Watermark Recovery Process

Input: improved double watermark image I_m
Output: restored robust watermark W_b
1: Split the improved double watermark image into R-, G- and B-channel images.
2: Extract the ciphertext robust watermark W_{eb} from channel B. In the encryption mode of scrambling and diffusion structure proposed in this paper, the attacker cannot crack the scrambling and diffusion keys and cannot extract and replace the robust watermark and steal copyright information. //Extraction and replacement attack test.
3: Change the DCT coefficient selected in the embedding process of channel B, so that the difference *Dif* is modified out of the original region when the embedding watermark bit is unchanged. //Destroying the attack test.
4: Combine the damaged B channel with R and G channels.
5: Extract the attacked ciphertext robust watermark W'_{eb} from the B channel of the combined image, and extract the improved fragile watermark W_{mc} and fragile watermark W_c successively from the G channel.
6: Synchronize the fragile watermark with the attack on the ciphertext robust watermark by XOR W'_{eb} and W_{mc}, and obtain the fragile watermark W'_c (Figure 13e), and use the extracted fragile watermark W_c (Figure 13d) to position tampering and obtain W_L (Figure 13f):

$$W_L = \text{bitxor}\big(W_c, (\text{bitxor}(W'_{eb}, W_{mc}))\big) \tag{21}$$

7: After confirming the tampering location result, recover the attacked ciphertext robust watermark (Figure 13g) and further decrypt it to obtain the recovery result of the original attacked robust watermark (Figure 13h) //Restoring the tampered robust watermark.
For i from 1 to 64 **Do**
  For j from 1 to 64 **Do**
    If ($W_c(i.j)$ ~= $W'_c(i.j)$)
      $W_L(i.j) \leftarrow$ abs(1− I_m)
Else
$W_L(i.j) \leftarrow W_{eb}$
End
End
End
8: $W_{eb} \leftarrow$ **Recover**(W_{eb},W_L) //Restoring the attacked ciphertext robust watermark
9: $W_b \leftarrow$ **Decryption**(W_{eb}) //Recovered robust watermark
10: The restored robust watermark is compared with the original robust watermark, and the result is that the all-zero image indicates that the two watermarks are identical
11: End procedure.

The simulation results (Figure 13) show that on the basis of the improved encryption method of robust watermark and the establishment of the correlation system between the fragile watermark and the robust watermark proposed in this paper, attackers cannot crack the key, extract and replace the robust watermark, indicating that the improved scheme can resist the extraction and replacement attack. Then, use the attack method proposed earlier to destroy the robust watermark. Using the tampered ciphertext robust watermark (b) to process the extracted improved fragile watermark (c), the obtained fragile watermark (e) can synchronize the attack signal of the ciphertext robust watermark (b) in channel B, so that their pixel changes are the same. Then, by comparing (d) and (e), we can know the specific pixel value that (b) changes, and then we can realize the tamper location of the robust watermark in the ciphertext in channel B. After accurately locating the tamper, first recover the attacked ciphertext robust watermark, and then perform the decryption operation to achieve the full recovery of the original tampered robust watermark. The full black image (i) shows that the restored robust watermark (h) is the same as the original robust watermark, which shows that the improvement measures we have given are also effective against destructive attacks. It can not only synchronize the signal of the fragile watermark and the encrypted robust watermark in the attacked channel B but also achieves the full recovery of the original tampered robust watermark. This is the uniqueness of our

proposed system to establish the correlation between the fragile watermark and the robust watermark. The tamper location and recovery performance of the improved scheme when the robust watermark is attacked is shown in the next section.

4.3. Performance Test

In order to evaluate the security and effectiveness of the improved watermarking scheme, we have carried out experimental tests in the aspects of perception quality, robustness, vulnerability, tamper location and recovery. The carrier image and watermark image used are shown in Figure 14; six carrier images are color standard test chart with the size of 512 × 512, namely 'Baboon', 'Sailboat',' Yacht',' Pepper', 'Tiffany', and 'Lena.' The robust watermark and fragile watermark are binary flag watermarks with the size of 64 × 64.

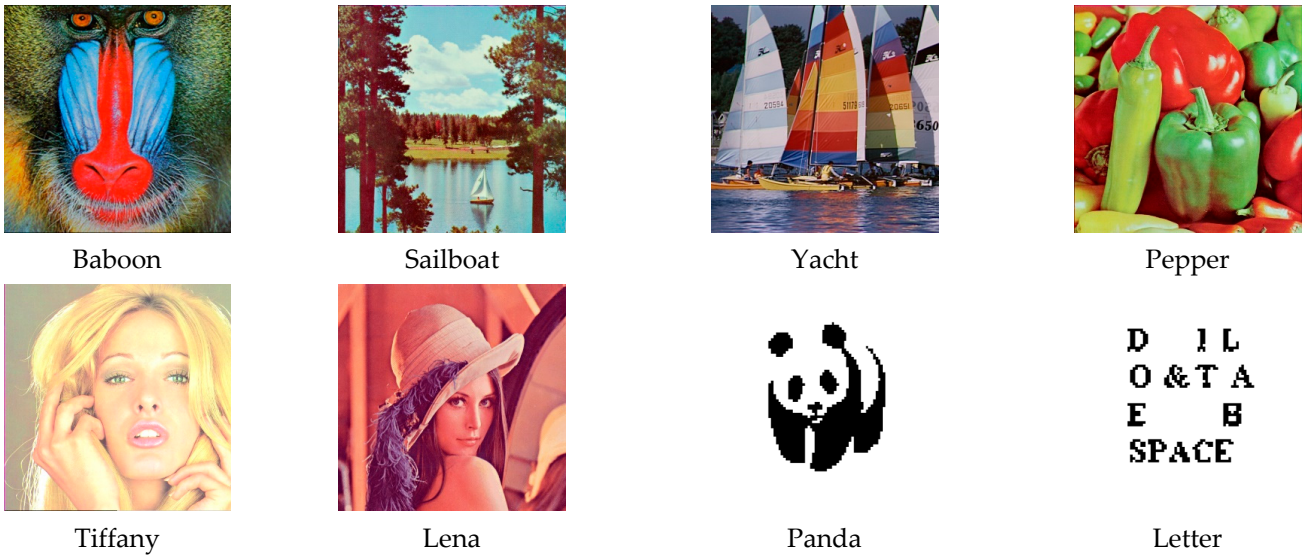


Figure 14. Six color test images in size 512 × 512 and two 64 × 64 size flag watermarks.

4.3.1. Perceived Quality Analysis

The perceptual quality of digital watermarking, also known as imperceptibility, requires the watermark to be imperceptible and without causing obvious degradation of the carrier image after embedding the digital content, thus reducing the attack intention of the attacker. Peak Signal-to-Noise Ratio (PSNR) and Structure Similarity Index (SSIM) are usually used to evaluate the imperceptibility of watermarks. PSNR is defined as follows:

$$\begin{aligned} \text{PSNR} &= 10 \times \log_{10}\left(\frac{255\times 255}{\text{MSE}}\right) \\ \text{MSE} &= \frac{1}{M\times N}\sum_{x=1}^{M-1}\sum_{y=1}^{N-1}\left(I_O(x,y) - I_W(x,y)\right)^2 \end{aligned} \tag{22}$$

MSE is the mean square error value calculated from the pixel value of the original carrier image $I_O(x,y)$ and the watermark image $I_W(x,y)$. The PSNR standard value is within 30–50 dB. The larger the value, the higher the image similarity. As an index to measure the similarity between the original carrier image $I_O(x,y)$ and the watermark image $I_W(x,y)$, SSIM is calculated as follows:

$$\text{SSIM}(I_O, I_W) = \frac{(2\mu_{I_O}\mu_{I_W} + c_1)(2\sigma_{I_O I_W} + c_2)}{(\mu_{I_O}^2 + \mu_{I_W}^2 + c_1)(\sigma_{I_O}^2 + \sigma_{I_W}^2 + c_2)} \tag{23}$$

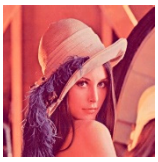
where μ_{I_O} and μ_{I_W} are the average values of image $I_O(x,y)$ and $I_W(x,y)$, $\mu_{I_O}^2$ and $\mu_{I_W}^2$ are the variances of image $I_O(x,y)$ and $I_W(x,y)$, $\sigma_{I_O I_W}$ are the covariance of $I_O(x,y)$ and $I_W(x,y)$, $c_1 = (k_1 L)^2$ and $c_2 = (k_2 L)^2$ are constants used to maintain stability, where

$L(L = 2^B - 1$, B is the bit depth) is the dynamic range of pixel values, typically taken as 255, k_1 is generally 0.01, and k_2 is 0.03. The SSIM value range is [0, 1]. The closer the value is to 1, the smaller the difference between the watermark image and the original image is, that is, the better the image quality is. Table 1 shows the PSNR and SSIM values of the double watermark images generated by different test color images. The results show that the PSNR values of various test color images are all higher than 41 dB, and the average SSIM value is higher than 0.99, indicating that the double watermark images generated by the improved scheme have a high similarity with the original test color images, and the double watermark image quality is very high. It fully shows that the double watermark images generated by the improved scheme have good imperceptibility. Table 2 shows the comparison results of the PSNR and SSIM values of the improved scheme and the original scheme using the test color images 'Baboon' and 'Lena' to generate dual watermark images. It can be seen that the SSIM values of the improved scheme and the original scheme are both close to 1, while the PSNR values of the improved scheme are slightly higher than the original scheme. Because the embedded watermark images are different, and there are differences also in the quality of watermark images, it is reasonable for the PSNR value to fluctuate in a small range. In general, the improvement scheme we proposed has a good performance in terms of perceived quality.

Table 1. PSNR and SSIM values of different dual watermarked images.

Dual Watermarked Images	PSNR (dB)	SSIM
Baboon	41.3705	0.9953
Sailboat	42.0679	0.9928
Yacht	42.2113	0.9931
Pepper	42.2385	0.9875
Tiffany	41.8367	0.9918
Lena	42.1729	0.9964

Table 2. Comparison results of PSNR and SSIM between the improved scheme and the original scheme [35].

Schemes \ Dual Watermarked Images / Data Indicators			
Original	PSNR (dB)	41.11	42.03
	SSIM	0.99	0.99
Improved	PSNR (dB)	41.37	42.17
	SSIM	0.99	0.99

4.3.2. Robustness and Vulnerability Analysis

The robustness and vulnerability of the watermarking scheme are the key performance measures for copyright protection and content authentication. The bit error rate (BER) and normalized cross-correlation coefficient (NCC) are used to measure the robustness and vulnerability of the improved scheme and are defined as follows:

$$\begin{aligned} \mathrm{NCC} &= \frac{\sum_{i=0}^{HW} I_{\mathrm{Wi}} I'_{\mathrm{Wi}}}{\sqrt{\sum_{\mathrm{i}=0}^{HW-1} I_{\mathrm{Wi}}^2 \sum_{i=0}^{HW-1} I_{\mathrm{Wi}}^2}} \\ \mathrm{BER} &= \frac{num\ of\ (F'_{bi} \neq F_{bi})}{length\ of\ F_b} \times 100\% \end{aligned} \tag{24}$$

where I_{Wi} and I'_{Wi} represents the ith bit of the original embedded watermark and extracted watermark. Under various attacks, the NCC value should be close to or equal to 1, and the BER value should be close to or equal to 0, indicating that the robustness is excellent. On the contrary, the NCC value is as small as possible, and the BER value is as high as possible, which indicates that the watermark has high vulnerability and can accurately locate the tamper.

We tested the improved double watermark image using multiple single-image processing and geometric attacks and their hybrid attacks in the original scheme [35]. Figure 15 shows the impact of these intentional or unintentional attacks on the double watermark image and the extracted robust watermark and fragile watermark. Table 3 shows the NCC and BER test values of the robust watermark and fragile watermark extracted under the condition of using multiple test color images. The results show that the extracted robust watermark can still be recognized under these attacks, the NCC value is close to 1, and the BER value is lower than 8% in all attack forms except for the attack of cropping lower half (50%). In fact, a bit error rate in the range of 10% to 20% is also acceptable. Experimental results indicate that the improved scheme has a high level of robustness, while the fragile watermark has been completely destroyed, indicating that intentional or unintentional tampering can be detected.

Table 3. Performance test of robust watermark and fragile watermark extracted under single attack.

Attacks	Watermark	Baboon		Sailboat		Pepper		Lena	
		BER	NCC	BER	NCC	BER	NCC	BER	NCC
Crop upper left (25%)	Robust	7.32	0.99	7.68	0.99	6.89	1	7.35	0.99
	Fragile	22.70	0.71	21.86	0.65	21.52	0.71	23.56	0.75
Crop lower right (25%)	Robust	7.34	0.99	7.65	0.98	6.88	0.99	7.33	0.99
	Fragile	22.35	0.70	23.58	0.67	21.68	0.65	22.26	0.74
Crop lower half (50%)	Robust	14.04	0.98	15.37	0.99	13.81	0.99	14.71	0.98
	Fragile	52.06	0.41	53.74	0.35	47.36	0.43	49.83	0.39
Median filter [3, 3]	Robust	1.92	0.99	1.67	0.99	1.26	0.97	2.51	0.97
	Fragile	48.31	0.48	45.29	0.42	46.85	0.51	47.37	0.54
Low pass filter	Robust	4.05	0.99	3.17	0.96	2.73	0.98	3.27	0.98
	Fragile	48.16	0.45	48.30	0.41	48.27	0.47	49.15	0.48
Average filter [3, 3]	Robust	2.89	0.97	4.48	0.98	2.61	0.99	3.23	0.99
	Fragile	48.93	0.51	47.25	0.49	50.10	0.53	46.91	0.50
Weiner filter [3, 3]	Robust	2.16	0.98	1.81	0.99	0.79	0.99	1.59	0.96
	Fragile	48.80	0.50	51.38	0.52	49.32	0.46	50.12	0.51
S & P noise (0.01)	Robust	5.70	0.98	5.44	0.95	6.14	0.98	5.13	0.99
	Fragile	39.28	0.73	36.06	0.68	38.42	0.74	35.84	0.71
Gaussian noise (0.001)	Robust	0.09	0.99	0.15	0.97	0.39	0.96	0.08	0.99
	Fragile	47.92	0.47	48.73	0.49	49.55	0.38	50.37	0.45
Speckle noise (0.01)	Robust	3.24	0.99	3.58	0.99	2.42	0.99	1.73	0.97
	Fragile	49.52	0.51	49.71	0.50	48.61	0.52	48.74	0.50
Poisson noise (0.01)	Robust	3.19	0.96	2.68	0.97	2.71	0.98	3.25	0.96
	Fragile	48.92	0.49	49.63	0.47	49.85	0.56	49.25	0.51

In addition, we also compared the improved scheme with the original scheme [35] and the literature [39], taking the 'Lena' color image of size 512 × 512 as the test standard, and obtained the BER and NCC test results of different schemes under various attack modes, as shown in Table 4.

It can be seen from Table 4 that the BER and NCC values obtained by the improved scheme are still better than the scheme [39] and the original scheme [35] under most attacks. Considering that watermarking is likely to suffer more than one intentional or unintentional attack in real application scenarios, in addition to the single attack, we also use the mixed attacks in the original scheme to evaluate the improved scheme, and the results are shown in Table 5. It can be seen that the BER and NCC values of the improved scheme are close

to the results of the original scheme in a small range under the hybrid attack composed of digital image processing operations, such as noise, filtering, cutting and rotation, etc. Obviously, the improved scheme also works well against hybrid attacks.

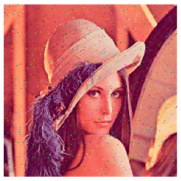
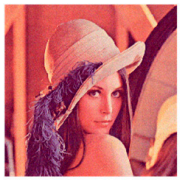
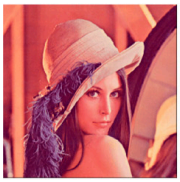
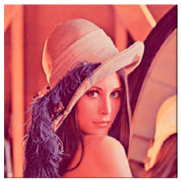
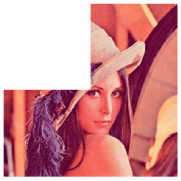
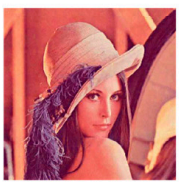
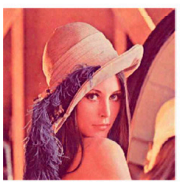
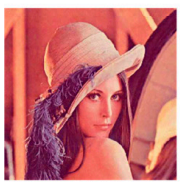
Attacks	Salt and pepper (0.01)	Gaussian noise (0.001)	Median filtering	Low Pass Filtering	Cropping (25%)
Attacked images					
Extracted watermarks	 	 	 	 	 
Attacks	JPEG Compression (40)	Rotation (15)	Resize down (50%)	Resize up (400%)	Sharpening
Attacked images					
Extracted watermarks	 	 	 	 	 

Figure 15. Double watermark image and extracted watermark image under multiple attacks.

Table 4. Comparison of the improved scheme with the original scheme [35] and scheme [39] in multiple attack modes.

Attack Type	[39]		[35]		Improved	
	BER	NCC	BER	NCC	BER	NCC
No attack	0	1	0	1	0	1
Salt and pepper (0.01)	0.0391	0.97	0.0550	0.96	0. 0513	0.9863
Poisson	0.1270	0.9713	0.0028	0.996	0.0121	0.9742
Speckle (0.01)	0.1211	0.91	0.0161	0.97	0.0173	0.9659
Average filter	0.0771	0.9451	0.0417	0.96	0.0350	0.9736
Gaussian LPF	0.0010	0.9993	0.0024	1	0.0021	1
Sharpening	0.0781	0.9455	0.066	0.975	0.0729	0.9783
Cropping (25%)	0.2500	0.7500	0.0166	1	0.1738	1
Cropping (50%)	0.5000	0. 5000	0.0753	0.998	0.0642	0.9937
Cropping (75%)	0.7500	0. 2500	0.073	0.989	0.0969	0.9841
LSB reset (1 or 2)	0	1	0	1	0	1
LSB reset (1–3)	0.0117	1	0	1	0	1
LSB reset (1–4)	0.1104	0.9205	0.0567	0.974	0.0548	0.9582
Resize (50%)	0.0518	0.9633	0.001	0.99	0.0232	0.9985

Table 5. Performance analysis of improved scheme and original scheme under mixed attack.

Combined Attacks	Watermark	[35]		Improved	
		BER	NCC	BER	NCC
S & P Noise (0.01) + Gaussian noise (0.001)	Robust	6.83	0.97	6.54	0.98
	Fragile	49.65	0.49	50.77	0.55
S & P Noise (0.01) + MF (3 × 3)	Robust	2.39	0.98	2.18	0.98
	Fragile	50.65	0.52	50.82	0.49
Rotation (10) + cropping (25%)	Robust	10.13	0.96	10.52	0.97
	Fragile	60.50	0.54	58.37	0.40
S & P Noise (0.01) + Crop (25%)	Robust	11.66	0.96	12.81	0.96
	Fragile	59.62	0.55	61.13	0.47
Scaling (400%) + Rotation (10)	Robust	3.61	0.98	3.37	0.98
	Fragile	61.02	0.48	63.02	0.39
S & P Noise (0.01) + MF (3 × 3) + Sharpening	Robust	5.85	0.98	5.73	0.98
	Fragile	60.19	0.49	59.42	0.46
Crop (25%) + Rotation (10) + Sharpening	Robust	12.35	0.95	14.04	0.97
	Fragile	61.45	0.35	60.29	0.41
Crop (25%) + Rotation (10) + Scaling (400%)	Robust	12.43	0.95	11.79	0.97
	Fragile	62.50	0.54	61.98	0.43

4.3.3. Tamper Detection and Recovery

Generally, attackers will replace part of the image information or add other different information in order to tamper with the copyright information. In order to test the tamper location and robust watermark recovery effect of the improved scheme, we use several attack methods in the original scheme to test the attacks on the dual watermark image and the B-channel image embedded with the robust watermark in the improved scheme. The fragile watermark is extracted to test the tampering location effect, and the robust watermark recovered after the attack is extracted to test the recovery effect, as shown in Figure 16. It can be seen that the improved scheme proposed in this paper has high-precision tampering location ability and a good tampering recovery effect for the common attacks that robust watermarks may suffer in a single channel.

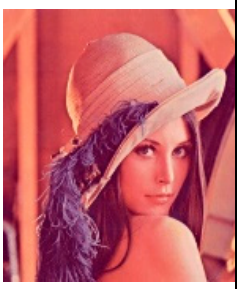
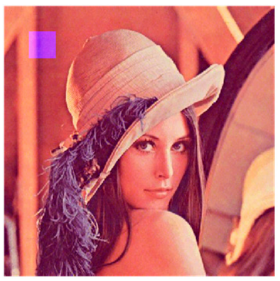
Attacks	Watermark Images	Attacked Watermark Images	Tamper Localization	Recovered Watermarks
Removal of 25% of a plane				
Replace of 25% of image with Pepper				
Add a scalar value to a square of pixels				
S & P noise (0.01)				

Figure 16. Tampering location and recovery robust watermark under some single attacks.

According to the above performance analysis results, the double watermark image generated by the improved scheme has an imperceptibility higher than 40 dB in many cases, and when subjected to various signal processing, geometric attacks, and hybrid attacks, the improved scheme shows excellent robustness and fragility. The experimental data obtained are almost close to or even higher than the original scheme and the literature [39] for comparison. It is worth mentioning that on the basis of overcoming the security vulnerabilities of the original scheme, the improved scheme enables the fragile

watermark to synchronize the attack signals of the robust watermark and utilizes the fragile watermark itself to achieve complete recovery of the tampered robust watermark. The experimental results in Figures 13 and 16 also verify the feasibility and effectiveness of the improved scheme.

5. Conclusions

This paper analyzes the security of a dual watermark technology scheme proposed by Hurrah et al. and gives some attacks. The main security flaws of the original scheme are that the double encryption scheme based on Arnold scrambling and sequence encryption for watermark is not safe, the scrambling key and equivalent encryption key are easily obtained, which makes the original scheme unable to resist the attack of the attacker to destroy, extract and replace the robust watermark. Moreover, there is no correlation between a robust watermark and a fragile watermark, and a fragile watermark cannot detect tampering under single-channel attack. In the improvement measures given in this paper, the double encryption method of a robust watermark has been changed, and the scrambling-diffusion encryption structure can ensure the sufficient security of the key. The improvement scheme also includes the establishment of an association system between a robust watermark and a fragile watermark, so that a fragile watermark can synchronize the attack signal of a ciphertext robust watermark and achieve accurate positioning of tampering and recovery of tampered areas. Simulation experiments and performance tests show that the attack strategy and improvement scheme proposed in this paper are feasible, safe and effective. The disadvantage of the improved scheme is that we can only achieve the recovery of tampered robust watermarks, and it is unable to correct the tampered areas of the dual watermark image. We intend to further facilitate recovery of the tampered areas of the located dual watermark image.

Author Contributions: M.L. proposed the research conceptualization and methodology; the software and improvement were performed by Y.Y.; the validation was performed by Y.Y.; writing of original draft, preparation, and editing were performed by Y.Y.; and the funding acquisition was performed by M.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Curriculum Reform Research Project of Teacher Education of Henan Province (Grant No. 2022-JSJYYB-003), the Science and Technology Research Project of Henan Province (Grant No. 212102210413), and the Key Program of the Higher Education Institutions of Henan Province (Grant No. 23A520009).

Data Availability Statement: The data that support the findings of this study are available on request from the corresponding author.

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

References

1. Samuel, A.; Sarfraz, M.I.; Haseeb, H.; Basalamah, S.; Ghafoor, A.A. Framework for Composition and Enforcement of Privacy-Aware and Context-Driven Authorization Mechanism for Multimedia Big Data. *IEEE Trans. Multim.* **2015**, *17*, 1484–1494. [CrossRef]
2. Jing, X.Y.; Yan, Z.; Pedrycz, W. Security Data Collection and Data Analytics in the Internet: A Survey. *IEEE Commun. Surv. Tutor.* **2019**, *21*, 586–618. [CrossRef]
3. Manazir, A.M.A.; Ihsan, A.; Bin Idris, M.Y.I.; Imran, M.; Shoaib, M. Countering Statistical Attacks in Cloud-Based Searchable Encryption. *Int. J. Parallel Program.* **2020**, *48*, 470–495.
4. Wu, L.; Du, X.; Fu, X. Security threats to mobile multimedia applications: Camera-based attacks on mobile phones. *IEEE Commun. Mag. Paper* **2014**, *52*, 80–87. [CrossRef]
5. Ma, S.; Zhang, T.; Wu, A.; Zhao, X. Lightweight and privacy-preserving data aggregation for mobile multimedia security. *IEEE Access* **2019**, *7*, 114131–114140. [CrossRef]
6. Chen, S.C. Digital Assets and Blockchain-Based Multimedia Data Management. *IEEE Multimed.* **2022**, *29*, 5–7. [CrossRef]
7. Ray, A.; Roy, S. Recent trends in image watermarking techniques for copyright protection: A survey. *Int. J. Multimed. Inf. Retr.* **2020**, *9*, 249–270. [CrossRef]
8. Guo, X.; Huang, D.Y.; Xu, L.T. A robust zero-watermarking scheme based on non-negative matrix factorization for audio protection. *PLoS ONE* **2022**, *17*, e0270579. [CrossRef]

9. Yuan, G.H.; Hao, Q. Digital Watermarking Secure Scheme for Remote Sensing Image Protection. *China Commun.* **2020**, *17*, 88–98. [CrossRef]
10. Pan, J.-S.; Sun, X.-X.; Chu, S.-C.; Abraham, A.; Yan, B. Digital watermarking with improved SMS applied for QR code. *Eng. Appl. Artif.* **2021**, *97*, 104049. [CrossRef]
11. Hurrah, N.N.; Khan, E.; Khan, U. CADEN: Cellular automata and DNA based secure framework for privacy preserving in IoT based healthcare. *J. Ambient Intell. Hum. Comput.* **2023**, *14*, 2631–2643. [CrossRef] [PubMed]
12. Kadian, P.; Shiafali, M.; Arora, N.A. Robust Digital Watermarking Techniques for Copyright Protection of Digital Data: A Survey. *Wirel. Pers. Commun.* **2013**, *118*, 3225–3249. [CrossRef]
13. Su, Q.; Niu, Y.; Zou, H.; Liu, X. A blind dual color images watermarking based on singular value decomposition. *Appl. Math. Comput.* **2013**, *219*, 8455–8458. [CrossRef]
14. Sreenivas, K.; Kamkshi Prasad, V. Fragile watermarking schemes for image authentication: A survey. *Int. J. Mach. Learn. Cybern.* **2018**, *9*, 1193–1218. [CrossRef]
15. Liu, X.L.; Lin, C.C.; Yuan, S.M. Blind Dual Watermarking for Color Images' Authentication and Copyright Protection. *IEEE Trans. Circuits Syst. Video Technol.* **2018**, *28*, 1047–1055. [CrossRef]
16. Swaraja, K.; Meenakshi, K.; Kora, P. An optimized blind dual medical image watermarking framework for tamper localization and content authentication in secured telemedicine. *Biomed. Signal Process. Control* **2020**, *55*, 101665.
17. Ahmadi, S.; Bagheri, B.; Zhang, G.; Rabbani, M.; Boukela, L.; Jelodar, H. An intelligent and blind dual color image watermarking for authentication and copyright protection. *Appl. Intell.* **2021**, *51*, 1701–1732. [CrossRef]
18. Kamili, A.; Hurrah, N.N.; Parah, S.A.; Bhat, G.M.; Muhammad, K. DWFCAT: Dual Watermarking Framework for Industrial Image Authentication and Tamper Localization. *IEEE Trans. Ind. Inform.* **2021**, *17*, 5108–5117. [CrossRef]
19. Al-Otum, H.M.; Ellubani, A.A.A. Secure and effective color image tampering detection and self-restoration using a dual watermarking approach. *Optik* **2022**, *262*, 169280. [CrossRef]
20. Sviatolsav, V.; Shelby, P.; Thierry, P. Attacks on Digital Watermarks: Classification, Estimation-Based Attacks, and Benchmarks. *IEEE Commun. Mag.* **2001**, *39*, 118–126.
21. Cayre, F.; Fontaine, C.; Furon, T. Watermarking security: Theory and practice. *IEEE Trans. Signal Process.* **2005**, *53*, 3976–3987. [CrossRef]
22. Teng, L.; Wang, X.Y.; Wang, X.K. Cryptanalysis and improvement of a chaotic system based fragile watermarking scheme. *AEU-Int. J. Electron. Commun.* **2013**, *67*, 540–547. [CrossRef]
23. Botta, M.; Cavagnino, D.; Pomponiu, V. A successful attack and revision of a chaotic system based fragile watermarking scheme for image tamper detection. *AEU-Int. J. Electron. Commun.* **2015**, *69*, 242–245. [CrossRef]
24. Rawat, S.; Raman, B. A chaotic system based fragile watermarking scheme for image tamper detection. *AEU-Int. J. Electron. Commun.* **2011**, *65*, 840–847. [CrossRef]
25. Li, M.; Zhang, J.H.; Wen, W.Y. Cryptanalysis and improvement of a binary watermark-based copyright protection scheme for remote sensing images. *Optik* **2014**, *125*, 7231–7234. [CrossRef]
26. Zhu, P.; Jia, F.; Zhang, J.L. A copyright protection watermarking algorithm for remote sensing image based on binary image watermark. *Optik* **2013**, *124*, 4177–4181. [CrossRef]
27. Nan, H.; Fang, B.; Yang, W.B.; Qian, J.Y.; Li, M.; Liu, Y.; Zhang, Y.S. Cryptanalysis and Improvement of the Robust and Blind Watermarking Scheme for Dual Color Image. *Math. Probl. Eng.* **2015**, *2015*, 526174. [CrossRef]
28. Su, Q.T.; Niu, Y.G.; Liu, X.X.; Yao, T. A novel blind digital watermarking algorithm for embedding color image into color image. *Optik* **2013**, *124*, 3254–3259. [CrossRef]
29. Niyishaka, P.; Bhagvati, C. Image splicing detection technique based on Illumination-Reflectance model and LBP. *Multim. Tools Appl.* **2020**, *80*, 2161–2175. [CrossRef]
30. Sivasubramanian, N.; Konganathan, G. A novel semi fragile watermarking technique for tamper detection and recovery using IWT and DCT. *Computing* **2020**, *102*, 1365–1384. [CrossRef]
31. Benrhouma, O. Cryptanalysis and improvement of a semi-fragile watermarking technique for tamper detection and recovery. *Multim. Tools Appl.* **2022**. [CrossRef]
32. Dadkhah, S.; Abd, M.A.; Hori, Y.; Ella, H.A.; Sadeghi, S. An effective SVD-based image tampering detection and self-recovery using active watermarking. *Signal Process. Image Commun.* **2014**, *29*, 1197–1210. [CrossRef]
33. Benrhouma, O.; Hermassi, H.; Belghith, S. Security analysis and improvement of an active watermarking system for image tampering detection using a self-recovery scheme. *Multimed. Tools Appl.* **2016**, *76*, 21133–21156. [CrossRef]
34. Swaraja, K.; Meenakshi, K.; Kora, P. Hierarchical multilevel framework using RDWT-QR optimized watermarking in telemedicine. *Biomed. Signal Process. Control* **2021**, *68*, 10268.
35. Hurrah, N.N.; Parah, S.A.; Loan, N.A.; Sheikh, J.A.; Elhoseny, M.; Muhammad, K. Dual watermarking framework for privacy protection and content authentication of multimedia. *Future Gener. Comput. Syst.* **2019**, *94*, 654–673. [CrossRef]
36. Lu, Z.M.; Liao, X.W. Counterfeiting attacks on two robust watermarking schemes. *Int. J. Innov. Comput. Inf. Control* **2006**, *2*, 841–848.
37. Kerckhofs, A. La cryptographie militaire. *J. Sci. Militaires* **1883**, *9*, 161–191.

38. Li, M.; Xiao, D.; Zhang, Y.S. Attack and Improvement of the Fidelity Preserved Fragile Watermarking of Digital Images. *Arab. J. Sci. Eng.* **2016**, *41*, 941–950. [CrossRef]
39. Abraham, J.; Paul, V. An imperceptible spatial domain color image watermarking scheme. *J. King Saud. Univ. Comput. Inf.* **2019**, *31*, 125–133. [CrossRef]