Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

TPE-H2MWD: an exact thumbnail preserving encryption scheme with hidden Markov model and weighted diffusion*

Xiuli CHAI[1,2], Xiuhui CHEN[1], Yakun MA[1], Fang ZUO[‡3], Zhihua GAN[‡2,3], Yushu ZHANG[4]

[1]*School of Artificial Intelligence, Henan Engineering Research Center for Industrial Internet of Things, Henan University, Zhengzhou 450046, China*

[2]*Henan Key Laboratory of Cyberspace Situation Awareness, Zhengzhou 450001, China*

[3]*School of Software, Intelligent Data Processing Engineering Research Center of Henan Province, Institute of Intelligent Network System, Henan University, Kaifeng 475004, China*

[4]*College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China*

E-mail: chaixiuli@henu.edu.cn; 2923105987@qq.com; 1060734169@qq.com; zuofang@henu.edu.cn; gzh@henu.edu.cn; yushu@nuaa.edu.cn

Received Oct. 21, 2022; Revision accepted Jan. 5, 2023; Crosschecked July 24, 2023

Abstract: With the substantial increase in image transmission, the demand for image security is increasing. Noise-like images can be obtained by conventional encryption schemes, and although the security of the images can be guaranteed, the noise-like images cannot be directly previewed and retrieved. Based on the rank-then-encipher method, some researchers have designed a three-pixel exact thumbnail preserving encryption (TPE2) scheme, which can be applied to balance the security and availability of images, but this scheme has low encryption efficiency. In this paper, we introduce an efficient exact thumbnail preserving encryption scheme. First, blocking and bit-plane decomposition operations are performed on the plaintext image. The zigzag scrambling model is used to change the bit positions in the lower four bit planes. Subsequently, an operation is devised to permute the higher four bit planes, which is an extended application of the hidden Markov model. Finally, according to the difference in bit weights in each bit plane, a bit-level weighted diffusion rule is established to generate an encrypted image and still maintain the same sum of pixels within the block. Simulation results show that the proposed scheme improves the encryption efficiency and can guarantee the availability of images while protecting their privacy.

Key words: Hidden Markov model; Weighted diffusion; Balance between usability and privacy; Image encryption
https://doi.org/10.1631/FITEE.2200498 **CLC number:** TN918; TP391

1 Introduction

With the development of the Internet and cloud service platforms, people upload their pictures and videos to cloud service platforms (Beaver et al., 2010). With fewer restrictions on hardware and easier file sharing, it is possible for users to access uploaded files at will (He et al., 2010). However, images may also contain sensitive information such as identity, location, and circumstances (Fan, 2019), so people are more inclined to prevent others from viewing their private images.

There are two main methods of privacy leakage: inside attacks and external attacks (Ashiq, 2015).

‡ Corresponding authors

* Project supported by the Pre-research Project of Songshan Laboratory, China (No.YYJC012022011), the Postgraduate Education Reform and Quality Improvement Project of Henan Province, China (Nos. YJS2022JD26 and SYLAL2023020), the Postgraduate Education Innovation Training Base, China (No. SYLJD2022008), the Science and Technology Project of Henan Province, China (Nos. 232102210109 and 232102210096), and the Open Foundation of Henan Key Laboratory of Cyberspace Situation Awareness, China (No. HNTS2022019)

ORCID: Xiuli CHAI, https://orcid.org/0000-0002-1609-0624; Fang ZUO, https://orcid.org/0000-0001-5673-8870; Zhihua GAN, https://orcid.org/0000-0002-2372-2853

Inside attacks are the most easily ignored security issue. Inside attacks generally refer to malicious damage by insiders, collusion between insiders and outsiders, abuse of power by managers, and threats caused by natural disasters such as floods and earthquakes. Even righteous employees can inadvertently leak data due to weak security awareness or incorrect use of software. External attacks refer to the attacks on nodes inside the network by malicious nodes outside the network that have not been authenticated. With the explosion of information and the maturity of attack techniques, there are increasingly more incidents of illegal intrusion in the network, which brings numerous difficulties to network security protection.

Some scholars have proposed conventional image encryption schemes to generate noise-like images (Joshi et al., 2020; Zhu et al., 2020). This prevents attackers from stealing the content in the image, but the image owners cannot identify the original image's content through encrypted images. To balance the availability and security of encrypted images in the cloud, Wright et al. (2015) introduced two thumbnail preserving encryption (TPE) schemes, where the plaintext images and the encrypted images have the same thumbnails. In the generated encrypted images, users with a priori knowledge can quickly find the image they want, while those without a priori knowledge can recognize only the general outline of the images and cannot obtain more detailed information. A priori knowledge refers to the impressions left in the user's mind by viewing the image.

The TPE schemes presented by Wright et al. (2015) change only the position of the pixels within the block, and although the thumbnail of the plaintext image is accurately preserved, the attackers can obtain all the pixel values in the plaintext image. Some studies (Li et al., 2008; Jolfaei et al., 2016) have shown that only the scrambling operation in the encryption process cannot satisfy the security requirements. To improve the security of encrypted images, Marohn et al. (2017) proposed two approximate TPE schemes, where the thumbnails of encrypted images and those of plaintext images are similar but not identical. The first method works using dynamic range preserving encryption (DRPE), which retains only the minimum and maximum pixel values of the plaintext in each thumbnail block. The second approach is an extended application of the least significant bit (LSB) embedding technique in steganography, which reveals only the maximum pixel value and the average value of the plaintext in the blocks. However, the encrypted images still reveal more information than the plaintext image thumbnails and cannot be fully decrypted to the original images. Later, Zhang et al. (2022) proposed a high-fidelity TPE scheme (HF-TPE), which improves the perceptual quality of encrypted images. Decrypted images have a lower noise intensity and an upper limit of noise quantity.

Tajik et al. (2019) achieved the first exact TPE scheme with nonce-respecting (NR) security. To preserve accurate thumbnails of the plaintext images, the adopted strategy is to ensure that the sum of pixels within a single block of a single channel remains unchanged. They summarized and employed the rank-then-encipher method proposed by Bellare et al. (2009). That is, first calculate the sum $\boldsymbol{S}$ of a vector, compute the rank of the vector through the function, and then encrypt the rank. Finally, the encrypted rank is converted into an integer vector via the function. All the vectors with the same sum $\boldsymbol{S}$ are numbered, and the rank of the vector is the corresponding number. This scheme encrypts the image in groups of two pixels, but it uses Markov chains to demonstrate its security. The fewer the number of pixels in a group, the worse the chain state connectivity. Therefore, based on the rank-then-encipher approach, Zhao et al. (2021) proposed the TPE2 scheme for encrypting images, which extended the exact TPE scheme to a group of three pixels for the first time; however, the TPE2 scheme has the problem of low efficiency.

To solve the above problems, we propose an exact TPE algorithm based on a hidden Markov model and weighted diffusion (TPE-H2MWD). The encryption process takes the entire block as a unit while changing the position and value of the pixels and keeping the sum of the pixels constant. After multiple rounds of encryption, an encrypted image with the same thumbnail as the plaintext image is obtained, making it impossible for attackers to identify whether the encrypted image comes from the plaintext image or from another image with the same sum of pixels in the block (e.g., a mosaic image of the plaintext) while improving the encryption efficiency. The main contributions of this paper are as follows:

1. A new exact TPE-H2MWD is presented. Bit-level permutation and diffusion methods are used to realize the exact TPE directly, which improves the encryption efficiency.

2. A bit-level permutation method based on the hidden Markov model is presented. First, blocking and bit-plane decomposition operations are performed on the plaintext image to obtain eight bit planes, and then different scrambling operations are performed on different bit planes. This makes the permutation process more random while improving the security of the encryption scheme.

3. A weighted diffusion method is introduced under the condition that the sum of pixels is unchanged. This method modifies the bit values and is able to prevent the leakage of statistical information of the bits.

4. In the TPE-H2MWD, the thumbnails of the ciphertext images and those of the plaintext images are absolutely identical, and the encrypted images can be recovered completely by the decryption process. Simulation results show that the proposed scheme balances the security and usability of encrypted images. In addition, there is good retrieval performance for thumbnails of ciphertext images.

2 Fundamentals

In this section, the definition of TPE and the fundamental knowledge of the hidden Markov model are presented.

2.1 Thumbnail preserving encryption

TPE is a special type of format-preserving encryption, which is the encryption of plaintext images into ciphertext images with the same properties. The definition of TPE is provided below.

Definition 1 In the TPE, a collection of images $\mathcal{M}$ is selected. An encryption scheme is claimed to be Φ-preserving within the plaintext image group $\mathcal{M}$, if it satisfies the following criteria:

$$\mathrm{Enc}_K(T, M) \to C, \tag{1}$$

$$\mathrm{Dec}_K(T, C) \to M, \tag{2}$$

$$\Phi(C) = \Phi(M), \tag{3}$$

where the plaintext image $M \in \mathcal{M}$, with the key K, nonce T, and the plaintext image M as input and the ciphertext image C as output in the encryption process, while with the key K, nonce T, and the ciphertext image C as input and the plaintext image M as output in the decryption process, as well as the plaintext image and the corresponding encrypted image having the same dimensionality and thumbnail.

Definition 2 Let $\mathcal{F}_\Phi$ be the set of functions F: $\{0,1\}^* \times \mathcal{M} \to \mathcal{M}$, where F satisfies the Φ-preserving property, and $\Phi(F(T, M))=\Phi(M)$ is met for all T and M. If for any probabilistic polynomial time (PPT) oracle machine $\mathcal{A}$, the following condition is satisfied (Bellare et al., 2009):

$$\left| \Pr_{K \leftarrow \{0,1\}^\lambda} \left[\mathcal{A}^{\mathrm{Enc}_K(\cdot,\cdot)}(\lambda) = 1 \right] - \Pr_{F \leftarrow \mathcal{F}_\Phi} \left[\mathcal{A}^{F(\cdot,\cdot)}(\lambda) = 1 \right] \right| \le \mathrm{negl}(\lambda), \tag{4}$$

where negl(λ) is negligible in λ, the format-preserving encryption scheme is deemed to satisfy the pseudo random permutation (PRP) security.

Definition 3 Let $\mathcal{A}$ be the PPT oracle machine with two parameters. $\mathcal{A}$ is denoted as NR if $\mathcal{A}$ never takes the same first parameter (i.e., nonce T) to make two oracle calls. A format-preserving encryption scheme is considered NR-secure if it can satisfy Definition 2 under only the condition that Definition 3 is respected (Tajik et al., 2019).

2.2 Hidden Markov model

Hidden Markov models are the statistical models (Franzese and Iuliano, 2019) that are used to describe Markov processes with hidden uncertain parameters. There are three elements:

1. Visible random sequence. This is used to describe the physical quantity of interest, which changes over time.

2. Hidden state sequence. Each physical quantity at each time point corresponds to a state quantity, and each state quantity cannot be directly observed.

3. Relationships among variables. The initial state quantity, the relationship between the current hidden state quantity and the next hidden state quantity, and the relationship between the current hidden state quantity and the visible random variable, all these three relationships or variables are described in probabilistic terms.

The hidden Markov model is a kind of statistical model with quite common applications. The theory was proposed by Baum and Petrie (1966) and subsequently became an important direction of signal processing over a decade of development. At present, the hidden Markov model has also started to be widely applied to the fields of speech recognition (Srivastava et al., 2022), smart home applications (Youngblood and Cook, 2007), and behavior recognition (Hartmann et al., 2022; Xue and Liu, 2022), among which the application of the hidden Markov model in image processing is also relatively successful and effective.

In our scheme, the hidden state chains and the visible state chains are generated based on one bit plane, and then another bit plane is scrambled based on the generated sequences. The data in each bit plane are different, making the permutation process more random and not easy to predict.

3 Proposed TPE-H2MWD

After performing bit-plane decomposition on the plaintext image, one may obtain eight bit planes, as shown in Fig. 1. From this figure, we can see that the lower four bit planes contain less information about the plaintext image, and the higher four bit planes contain more information about the image.

Inspired by this, an exact TPE-H2MWD is proposed, and the specific algorithm framework is illustrated in Fig. 2. First, the plaintext image P ($M\times N\times 3$) is decomposed to obtain the red, green, and blue components, and then blocking and binary decompositions are performed on each component in turn to yield eight bit planes. The lower four bit planes are randomly zigzag scrambled, and then the scrambled lower four bit planes are used to generate two hidden state sequences and two visible state sequences. The higher four bit planes are scrambled with these sequences, and then, a weighted diffusion operation is performed between two adjacent bit planes. Finally, an encrypted image is obtained.

3.1 Zigzag model permutation (ZMP) for the lower four bit planes

The pixel block B of size $K\times K$ is binarily decomposed to obtain eight bit planes, denoted as B_8, B_7, B_6, B_5, B_4, B_3, B_2, and B_1. Zigzag scrambling operations are performed on the lower four bit planes B_1, B_2, B_3, and B_4 to obtain four new bit planes B_1', B_2', B_3', and B_4', respectively. Generally, there are eight types of zigzag scrambling models, as shown in Fig. 3, where the traversal rules of different models will be changed.

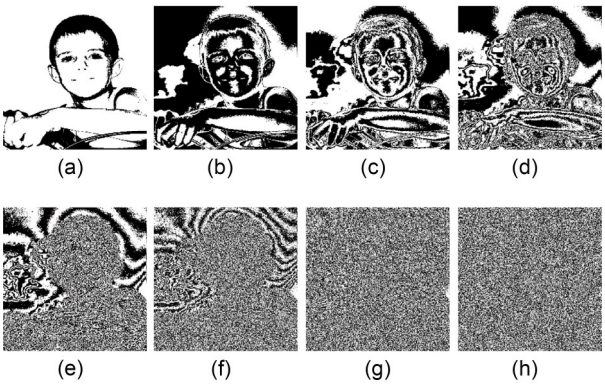

Fig. 1 Bit-plane decomposition: (a) 8th bit plane; (b) 7th bit plane; (c) 6th bit plane; (d) 5th bit plane; (e) 4th bit plane; (f) 3rd bit plane; (g) 2nd bit plane; (h) 1st bit plane

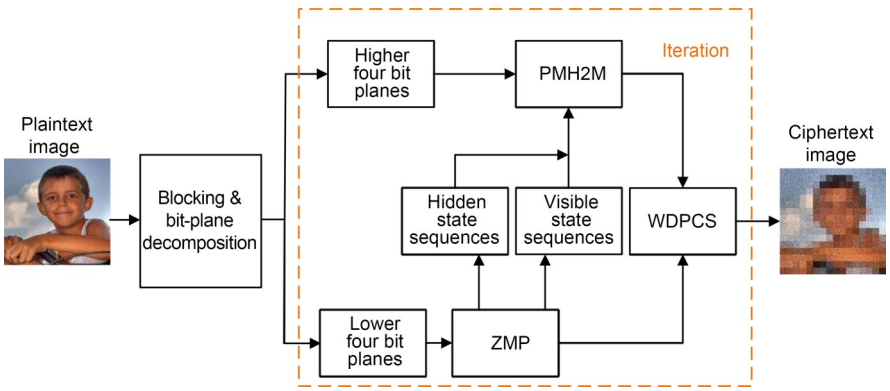

Fig. 2 Flowchart of the proposed scheme (ZMP: zigzag model permutation; PMH2M: permutation method based on hidden Markov model; WDPCS: weighted diffusion preserving constant sum)

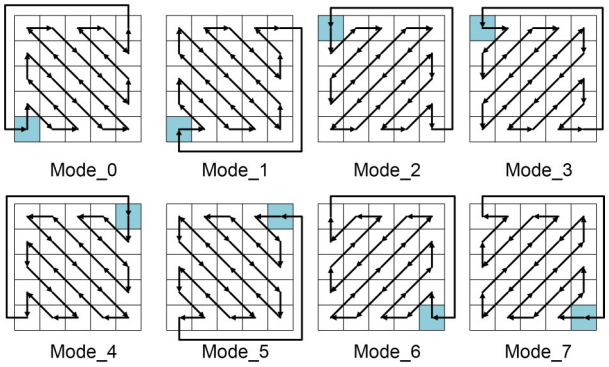

Fig. 3 Eight types of zigzag models

In this scheme, to improve the randomness of confusion, different zigzag models and starting positions are selected for different bit planes, and accordingly, the data in this bit plane are scrambled. The selected zigzag model is determined by ((count1−count0) % 8), and the starting position can be computed by ((count0 % K)+1, (count1 % K)+1), where count1 indicates the total number of bits with value 1 in the current bit plane, and count0 indicates the total number of bits with value 0 in the current bit plane.

3.2 Permutation method based on hidden Markov model (PMH2M) for the higher four bit planes

Four new bit planes B'_5, B'_6, B'_7, and B'_8 are generated by scrambling the four bit planes B_5, B_6, B_7, and B_8, respectively. The bits in B'_1, B'_2, B'_3, and B'_4 are employed to generate the hidden and visible state sequences. The data in the hidden state sequences are applied to determine bit populations, and bits in the bit population are to be involved in the shifting operation. The data in the visible state sequences are used to determine the distance that bits move. Specifically, the data in the lower plane B'_i are used to change the position of the bits in the higher plane B_j, and the detailed steps are as follows:

Step 1: Based on the bit values in the bit plane B_i', two hidden state sequences (Imp1, Imp2) and two visible state sequences (Vis1, Vis2) are created, and all of them have length K. The process of sequence generation is expressed by

$$\begin{cases} P_1 = t, \\ Q_1 = \begin{cases} K, & \text{if mod}\,(t+1, K) = 0, \\ \text{mod}\,(t+1, K), & \text{otherwise}, \end{cases} \\ R_1 = \begin{cases} K, & \text{if mod}\,(t+2, K) = 0, \\ \text{mod}\,(t+2, K), & \text{otherwise}, \end{cases} \end{cases} \tag{5}$$

where t is an integer, $t\in[1, K]$.

$$\begin{cases} \text{Imp1}_t = \text{mod}\left(\text{bin2dec}\left(\left(B'_i\right)_{1P_1}\left(B'_i\right)_{1Q_1}\left(B'_i\right)_{1R_1}\right), 4\right), \\ \text{Imp2}_t = \text{mod}\left(\text{bin2dec}\left(\left(B'_i\right)_{P_1 1}\left(B'_i\right)_{Q_1 1}\left(B'_i\right)_{R_1 1}\right), 4\right), \end{cases} \tag{6}$$

where Imp1_t and Imp2_t are the t^{th} elements in the hidden state sequences Imp1 and Imp2, respectively.

$$\begin{cases} \text{Vis1}_t = \text{mod}\left(\sum_{e=1}^{K}\left(\left(B'_i\right)_{eP_1} + \left(B'_i\right)_{eQ_1} + \left(B'_i\right)_{eR_1}\right), K\right), \\ \text{Vis2}_t = \text{mod}\left(\sum_{e=1}^{K}\left(\left(B'_i\right)_{P_1 e} + \left(B'_i\right)_{Q_1 e} + \left(B'_i\right)_{R_1 e}\right), K\right), \end{cases} \tag{7}$$

where Vis1_t and Vis2_t are the t^{th} elements in the visible state sequences Vis1 and Vis2, respectively.

To generate the sequence Imp1, the P_1^{th}, Q_1^{th}, and R_1^{th} bits of the first row in B'_i are taken, three bits are converted to a decimal number, and the remainder of the number divided by 4 is the t^{th} data Imp1_t. Similarly, to generate the sequence Imp2, the P_1^{th}, Q_1^{th}, and R_1^{th} bits of the first column in B'_i are selected, and then the t^{th} data Imp2_t are obtained.

To generate the sequence Vis1, the total number of bits 1 in the P_1^{th}, Q_1^{th}, and R_1^{th} columns of B'_i is taken, and the remainder of the total number divided by K is the t^{th} data Vis1_t. Similarly, to generate the sequence Vis2, the total number of bits 1 in the P_1^{th}, Q_1^{th}, and R_1^{th} rows of B'_i is obtained, and the remainder of the total number divided by K is the t^{th} data Vis2_t.

A complete example of generating sequences is shown in Fig. 4.

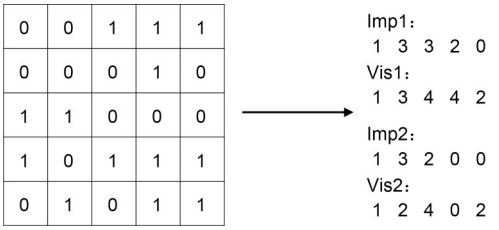

Fig. 4 Generating two hidden state sequences (Imp1, Imp2) and two visible state sequences (Vis1, Vis2)

Step 2: According to the data in Imp1 and Vis1, the bit populations and the moving distance of the

bits will be determined. In bit plane B_j, the first row of bits is used as a reference to change the position of the bits, and the new bit plane B_j' is obtained.

For bit populations in the minor and principal diagonal directions, circular downwards shift operations will be performed, and the moving distances are determined by the data in the corresponding visible state sequence. For bit populations in the row direction (as shown in Algorithm 1), bits in odd positions shift sequentially to the left, and bits in even positions shift sequentially to the right. For bit populations in the column direction, bits in odd positions shift successively upwards, and bits in even positions shift successively downwards. After moving the corresponding distance, if there are already data at the target location, continue moving one bit per time until a vacant location is found.

Algorithm 1 Confusion method of bit populations in the row direction

Input: bit sequence row1, shifting distance b.

Output: bit sequence row2.

```
n ← length(row1);
for i from 1 to n do
    if mod(i, 2) = 0 then
        r ← the first position in row2 that is not allocated a bit, starting with the b^th bit to the right of the i^th bit, circularly searching to the right;
    else
        r ← the first position in row2 that is not allocated a bit, starting with the b^th bit to the left of the i^th bit, circularly searching to the left;
    end if
    row2(r) ← row1(i);
end for
```

As shown in Fig. 5, the first data point in Imp1 is 1, and then the set of bits in the column where the first bit in the first row in the plane B_j is located is called the bit population. The first data point in the visible state sequence Vis1 is 1. When finding the target position in this bit population, it moves one bit first. The 1st bit is moved up cyclically by one bit, reaching the 5th position of the first column in B_{j1}; the 2nd bit is moved down cyclically by one bit, reaching the 3rd position of the first column in B_{j1}; the 3rd bit is moved up cyclically by one bit, reaching the 2nd position of the first column in B_{j1}; the 4th bit is moved down cyclically by one bit, reaching the 5th position of the first column, but there are already data in the 5th position, so it continues down one bit to the 1st position in the first column in B_{j1}; the 5th bit is moved up cyclically by one bit, reaching the 4th position of the first column in B_{j1} (the resulting bit population is outlined by the blue line in B_{j1}). The other operations are similar.

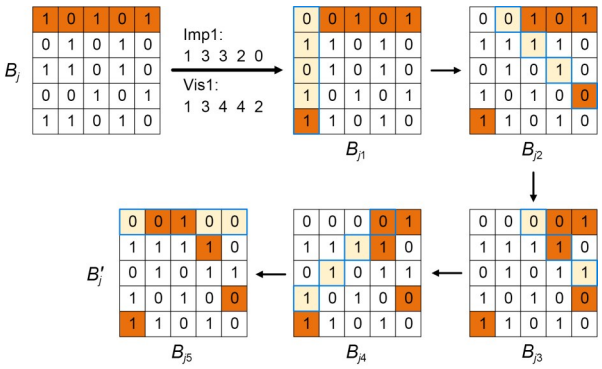

Fig. 5 Illustration of the shift of bits based on data in the hidden and visible state sequences (References to color refer to the online version of this figure)

Step 3: According to the data in Imp2 and Vis2, the bit populations and the moving distance of the bits will be determined. In bit plane B_j', the first column of bits is used as a reference to change the position of the bits, and the new bit plane B_j'' is obtained.

3.3 Weighted diffusion preserving constant sum (WDPCS)

A bit-level weighted diffusion scheme that maintains sum invariance is proposed using the different weights occupied by bits in various bit planes.

For eight bit planes, two adjacent bit planes are selected as a group, which can be divided into seven groups in total. Starting from the group containing the lowest bit plane, perform the following operations on each group successively: for two adjacent bit planes B_i and B_{i+1} with dimensions $K\times K$, the weights occupied by the bits in B_{i+1} are twice the weights of the bits in B_i. Each bit plane is divided into two equal parts on the left and right. If K is odd, the first $K-1$ columns are picked for the operation. Then, three bits are chosen from B_i and B_{i+1}, one bit is from the left part of B_{i+1}, and two bits correspond to the same position in the left and right parts of B_i. If the three bits are 0, 1, 1 (or 1, 0, 0), the bits can be converted

to 1, 0, 0 (or 0, 1, 1), and the sum of the values represented by all bits in B_i and B_{i+1} remains constant. Repeat this operation until all the bits in the left part of B_i have been selected. A concrete example is shown in Fig. 6.

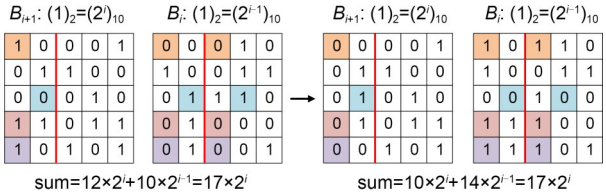

Fig. 6 Variation of bits in two adjacent bit planes (References to color refer to the online version of this figure)

In Fig. 6, the weight of each bit in B_{i+1} is 2^i, and the weight of each bit in B_i is 2^{i-1}. When the three selected bits are 1, 0, 0 (or 0, 1, 1), they are marked with the same color and modified to 0, 1, 1 (or 1, 0, 0). Before and after modification, the sum of the values represented by all bits in B_i and B_{i+1} is 17×2^i.

4 Simulations

In this section, a Helen dataset (Zhao et al., 2021) containing 500 face images with a size of 512×512 and a Holiday dataset (Jegou et al., 2008) containing 1491 holiday photos with a size of 512×512 are used to assess the empirical evaluation of the proposed scheme.

4.1 Analysis of the encryption effect

The encrypted images are obtained by iterating the encryption process for 10 rounds. As shown in Fig. 7, the block size can be adjusted to achieve a balance between the privacy and availability of the image. When K=0, the image indicates a plaintext image; when K reaches 8 and 16, many details are shown in the encrypted images, which is not secure enough; when the block size is 32×32, the encrypted image retains the rough information of the plaintext image for availability, and it also has a certain level of security.

4.2 Ciphertext image perception quality

The peak signal-to-noise ratio (PSNR) is an objective standard for evaluating image quality (Chai et al., 2022a). The larger the value is, the less distortion

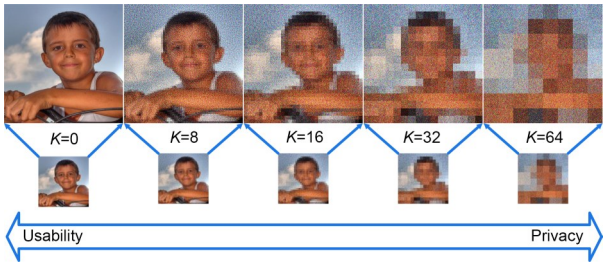

Fig. 7 TPE-H2MWD encrypted images with different block sizes (original, 8×8, 16×16, 32×32, and 64×64) and corresponding preserved thumbnails

of the images is. Structural similarity (SSIM) (Wu et al., 2022) is another metric for measuring the similarity of two images in terms of brightness, contrast, and structure, and the closer its value is to 1, the more similar the two images are.

Five hundred images from the Helen dataset are used for testing. The block sizes are set to 8×8, 16×16, and 32×32, corresponding to thumbnail sizes of 64×64, 32×32, and 16×16, respectively. That is, the pixel average value within each block corresponds to a value in the thumbnail. From the data in Tables 1 and 2, it can be seen that the mean value of PSNR for the TPE-H2MWD and TPE2 schemes reached +∞ and the mean value of SSIM reached 1, indicating that

Table 1 Peak signal-to-noise ratio (PSNR) values between 500 plaintext and encrypted image thumbnails

Algorithm	PSNR (dB)		
	8×8	16×16	32×32
TPE-LSB (1-bit) (Marohn et al., 2017)	20.045	20.532	21.347
TPE-LSB (2-bit)	29.545	30.393	31.889
TPE-LSB (3-bit)	39.145	40.675	43.327
TPE2 (Zhao et al., 2021)	+∞	+∞	+∞
TPE-H2MWD	+∞	+∞	+∞

Table 2 Structural similarity (SSIM) values between 500 plaintext and encrypted image thumbnails

Algorithm	SSIM		
	8×8	16×16	32×32
TPE-LSB (1-bit) (Marohn et al., 2017)	0.809	0.867	0.936
TPE-LSB (2-bit)	0.937	0.970	0.994
TPE-LSB (3-bit)	0.975	0.991	0.999
TPE2 (Zhao et al., 2021)	1	1	1
TPE-H2MWD	1	1	1

encrypted images completely preserve the thumbnails of the corresponding plaintext images.

4.3 Information entropy analysis

Information entropy is used to reflect the randomness of the information. Information entropy can be computed by (Mishra et al., 2021)

$$H(m)=\sum_{i=0}^{2^n-1}p(m_i)\log_2\frac{1}{p(m_i)}, \tag{8}$$

where $p(m_i)$ denotes the probability of occurrence of the pixel value m_i, and n refers to the bit depth of the image, which has 8-bit depth for grayscale images.

When the block size is 32×32, the plaintext image shown in Fig. 7 is encrypted with different schemes, and the information entropies of the obtained ciphertext images are listed in Table 3. It is clear that among the approximate TPE schemes, the information entropy obtained by the TPE-LSB (1-bit) scheme is the largest; in the exact TPE schemes, the information entropy obtained by the TPE2 scheme is slightly larger than the result in TPE-H2MWD, indicating that the ciphertext image in the TPE2 scheme has stronger randomness.

Table 3 Information entropy of plaintext and ciphertext images

Algorithm	Entropy		
	R	G	B
Plaintext image	7.366	7.561	7.351
TPE-LSB (1-bit) (Marohn et al., 2017)	7.992	7.992	7.974
TPE-LSB (2-bit)	7.919	7.989	7.958
TPE-LSB (3-bit)	7.920	7.989	7.958
TPE2 (Zhao et al., 2021)	7.929	7.982	7.929
TPE-H2MWD	7.925	7.973	7.907

4.4 Histogram analysis

In image processing, the pixel values of specific channels of an image are used as the x axis of the histogram, the number of occurrences of each value is used as the y axis, and the resulting histogram is the image histogram. The histogram of the image reflects the distribution density of image pixels. The images in Fig. 7 are selected for testing, and the histograms of the plaintext image and ciphertext images are shown in Fig. 8. The ciphertext image histogram contains some features of the plaintext image thumbnail, thus allowing effective use of the encrypted image (Chai et al., 2022b).

4.5 Adjacent pixel correlation analysis

The adjacent pixel correlation reflects the degree of correlation between pixel values at adjacent locations in the image, including mainly the correlation among horizontal pixels, vertical pixels, and diagonal pixels in the image. Here, we randomly select 5000 pairs of neighboring pixels (horizontal, vertical, and diagonal directions) from the plaintext and ciphertext images and then calculate the correlation coefficient by (Chen et al., 2021)

$$\begin{cases} E(x)=\dfrac{1}{N}\displaystyle\sum_{i=1}^{N}x_i, \\ D(x)=\dfrac{1}{N}\displaystyle\sum_{i=1}^{N}\big(x_i-E(x)\big)^2, \\ \mathrm{CC}=\dfrac{\dfrac{1}{N}\displaystyle\sum_{j=1}^{N}\big(x_j-E(x)\big)\big(y_j-E(y)\big)}{\sqrt{D(x)D(y)}}, \end{cases} \tag{9}$$

where x and y refer to the values of two adjacent pixels in the image, N refers to the pair of pixels selected from the image, and $E(x)$ and $D(x)$ refer to the mathematical expectation and variance of x, respectively, where the larger the correlation coefficient is, the greater the adjacent pixel correlation is, and vice versa.

The plaintext image shown in Fig. 7 is encrypted when the block size is 32×32, and the results are shown in Fig. 9. In addition, the adjacent pixel correlations of the encrypted images corresponding to the different encryption schemes are listed in Table 4. It can be concluded that compared to the approximate TPE scheme (TPE-LSB), the ciphertext image obtained by the TPE-H2MWD scheme has the lower adjacent pixel correlation; furthermore, compared to the exact TPE scheme (TPE2), our data still have significant advantages.

4.6 Running time

The time complexity of the TPE2 scheme is $O(mnd^3)$, where m and n are the dimensions of the plaintext image and d is the maximum value of pixels in the plaintext image, and the time complexity of

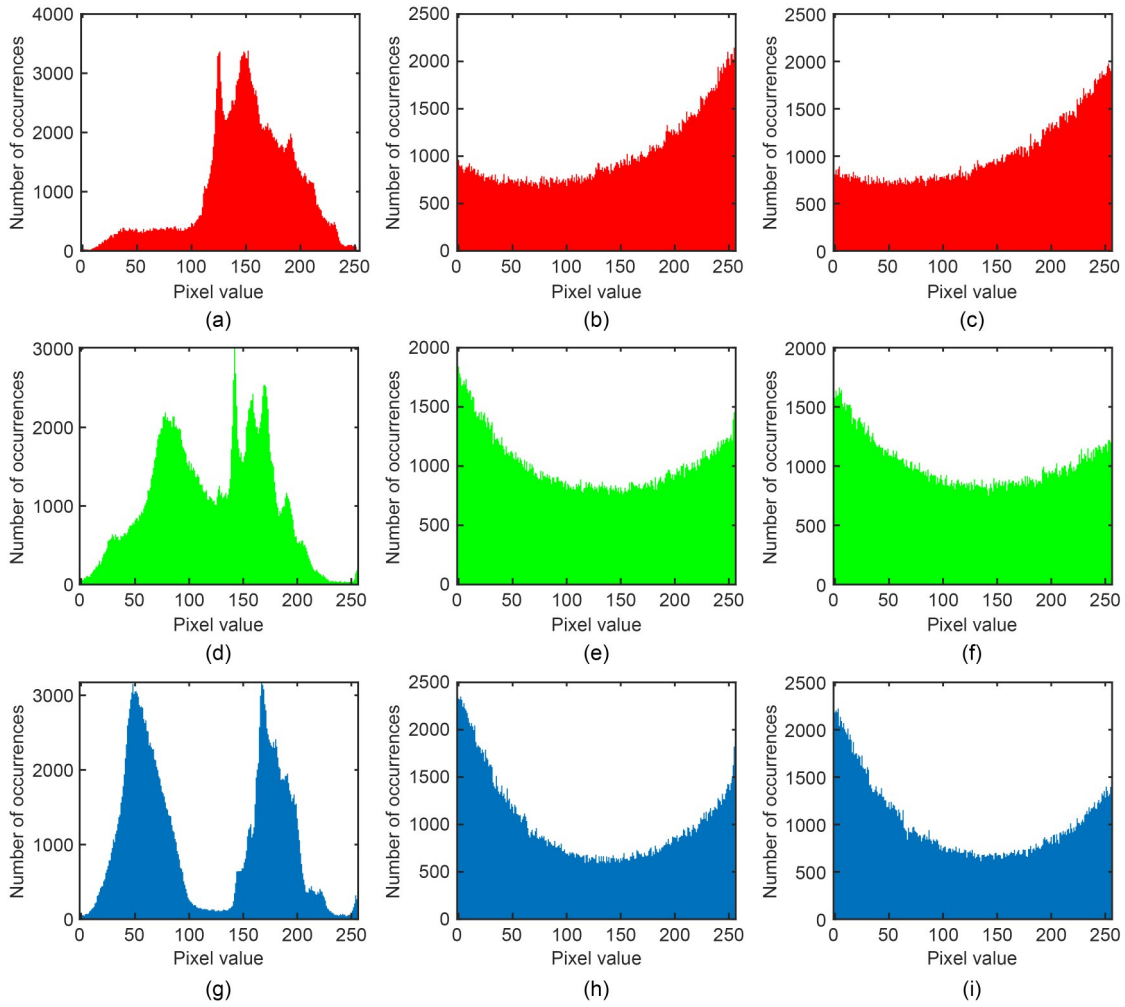

Fig. 8 Histogram in plaintext image and ciphertext images: (a)−(c) are histograms of the R component of the plaintext image and ciphertext images with *K*=8 and 32, respectively; (d)−(f) are histograms of the G component of the plaintext image and ciphertext images with *K*=8 and 32, respectively; (g)−(i) are histograms of the B component of the plaintext image and ciphertext images with *K*=8 and 32, respectively

the TPE-H2MWD scheme is $O(mn)$. Here, we intercept a subblock of a single channel in the color image for encryption, where the block size is 32×32. The data in Table 5 show that the encryption time of the exact TPE scheme (TPE2) grows rapidly with increasing subblock size and is longer than the running time in TPE-H2MWD. The approximate TPE scheme (TPE-LSB) has the shortest encryption time.

4.7 Image retrieval effect test

At present, content-based image retrieval is a mainstream retrieval method. Encrypted image thumbnails are obtained using in-block averaging, and then the color histogram feature and the depth feature based on VGG16 are extracted from the ciphertext image thumbnails. The cosine distance between feature vectors is used to calculate the similarity between images, and finally, the image retrieval accuracy is obtained. Encrypted image thumbnails from the Holiday dataset are tested, and the mean average precision (mAP) results are listed in Tables 6 and 7. The larger the mAP value is, the higher the retrieval accuracy is. Plaintext images have the highest retrieval accuracy. In addition, when the block size reaches 16×16, 32×32, and 64×64, the highest retrieval accuracy can be obtained if the thumbnail size is 32×32; when the block size reaches 8×8, the highest retrieval accuracy can be obtained if the thumbnail size is 64×64. Thus, users can choose the block size and thumbnail size according to their specific requirements.

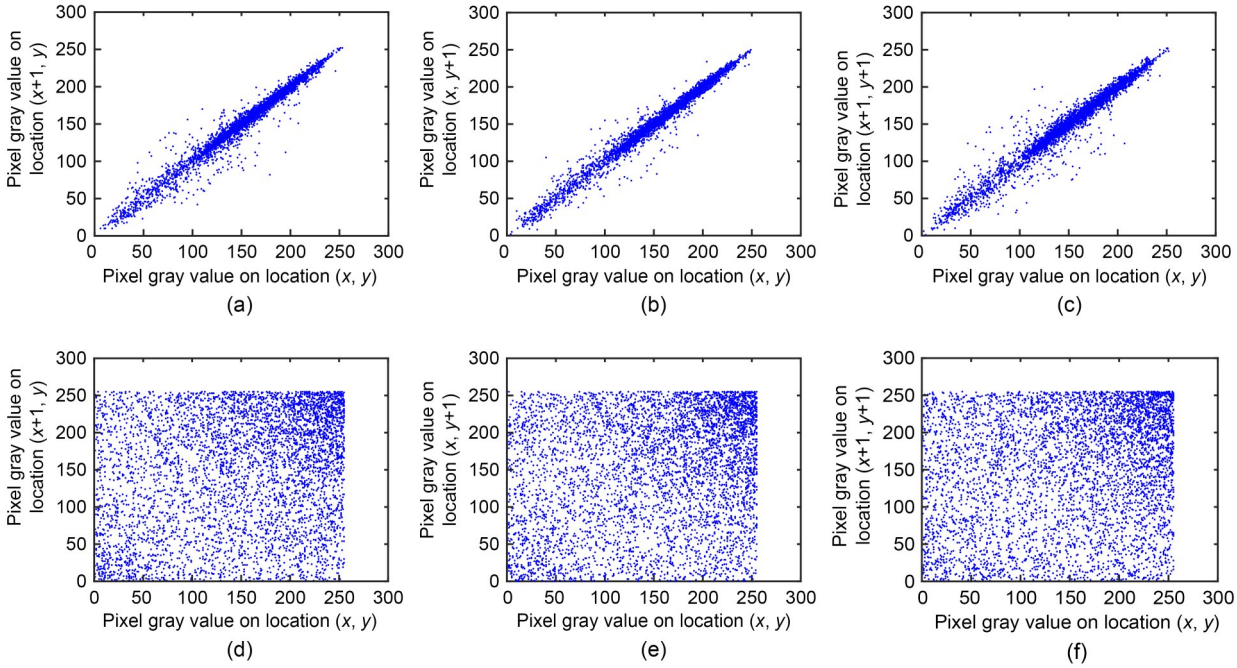

Fig. 9 Correlation analysis of the R channel in the image: (a)–(c) are the results of correlation of plaintext image in the horizontal, vertical, and diagonal directions, respectively; (d)–(f) are the results of correlation of encrypted image with K= 32 in the horizontal, vertical, and diagonal directions, respectively

Table 4 Correlation coefficients of plaintext and ciphertext images

Algorithm	Correlation coefficient								
	R			G			B		
	H	V	D	H	V	D	H	V	D
TPE-LSB (3-bit) (Marohn et al., 2017)	0.339	0.359	0.311	0.443	0.441	0.446	0.647	0.643	0.628
TPE2 (Zhao et al., 2021)	0.255	0.261	0.224	0.342	**0.328**	0.355	0.547	0.536	0.541
TPE-H2MWD	**0.209**	**0.209**	**0.220**	**0.297**	0.329	**0.305**	**0.510**	**0.503**	**0.485**

H: horizontal; V: vertical; D: diagonal. Best results are in bold

Table 5 Running time

Algorithm	Running time (s)		
	32×32	64×64	128×128
TPE-LSB (1-bit) (Marohn et al., 2017)	0.024	0.026	0.042
TPE-LSB (2-bit)	0.025	0.028	0.044
TPE-LSB (3-bit)	0.023	0.031	0.043
TPE2 (Zhao et al., 2021)	**6.788**	**25.707**	**101.351**
TPE-H2MWD	0.509	0.972	2.411

Best results are in bold

Table 6 Mean average precision (mAP) values of image retrieval on encrypted image thumbnails based on the color histogram features

Thumbnail size	mAP				
	Plaintext image	Block size			
		8×8	16×16	32×32	64×64
8×8	–	0.365	0.365	0.365	0.365
16×16	–	0.469	0.469	0.469	0.413
32×32	–	0.498	0.498	0.471	0.413
64×64	–	0.520	0.463	0.438	0.394
512×512	0.537	–	–	–	–

4.8 Security analysis

Most likely, the biggest threat to TPE algorithms is super resolution (SR) technology, which reconstructs low-resolution images into corresponding high-resolution images by specific algorithms. Fig. 10a shows an image with a resolution of 128×128, Fig. 10b corresponds to a thumbnail with a resolution of 32×32, and the results obtained by reconstructing the thumbnail with advanced SR algorithms FSRCNN (Dong et al., 2016) and DASR (Wang et al., 2021) are shown

Table 7 Mean average precision (mAP) values of image retrieval on encrypted image thumbnails based on VGG16 depth features

Thumbnail size	mAP				
	Plaintext image	Block size			
		8×8	16×16	32×32	64×64
8×8	–	0.226	0.226	0.226	0.226
16×16	–	0.287	0.287	0.287	0.209
32×32	–	0.485	0.485	0.297	0.223
64×64	–	0.571	0.456	0.293	0.210
512×512	0.674	–	–	–	–

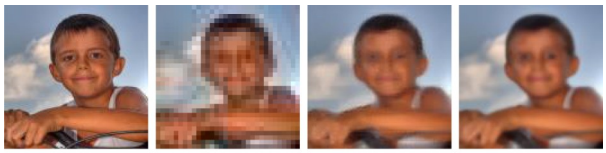

Fig. 10 Super resolution simulation results: (a) image with a resolution of 128×128; (b) a thumbnail with a resolution of 32×32; (c) results obtained by reconstructing the thumbnail with FSRCNN; (d) results obtained by reconstructing the thumbnail with DASR

in Figs. 10c and 10d, respectively. While the SR technique can enhance the visual quality of thumbnail images, the details of the reconstructed images and the original images differ, suggesting that the SR scheme restores only the approximate images but not the real ones.

4.9 Size expansion rate

In this subsection, we use the PNG images obtained from the Helen dataset to evaluate the expansion rate of the encrypted images. Suppose that the file size of the encrypted image is x and the file size of the plaintext image is y. Then, the encrypted image expansion rate is equal to x/y. For images in lossless PNG format, the redundancy of image information is used to accomplish lossless image compression. The redundancy of the plaintext image tends to be more than the redundancy of the encrypted image, resulting in a lower compression rate of the encrypted image than the plaintext image. As shown in Fig. 11, the average encrypted image expansion rates in different methods are demonstrated. In the TPE-H2MWD scheme, the cipher image expansion rate expands with increasing block size because the larger the block size is, the greater the number of available bit random

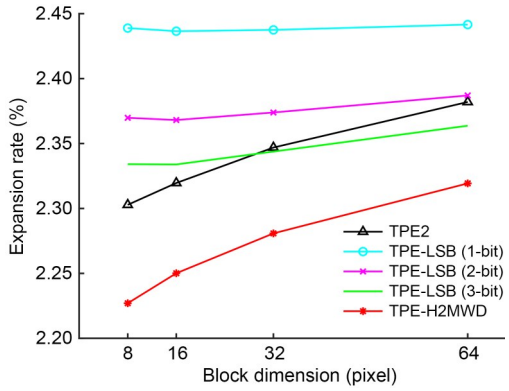

Fig. 11 Average size expansion rate of encrypted images

shift positions, the more severely the adjacent pixel correlation is broken, and the lower the compression efficiency.

5 Conclusions

In this paper, an exact TPE is implemented with hidden Markov model and weighted diffusion, and the specific encryption process and simulation results are given. The proposed TPE-H2MWD consists of two stages: bit-level permutation and weighted diffusion. In the scrambling phase, the zigzag models are applied in the lower four bit planes, and the hidden Markov model is used in the higher four bit planes. In the diffusion phase, the sum of the pixel values is kept constant. Simulation results show that the TPE-H2MWD can effectively balance the security and availability of images, where illegal users cannot obtain detailed information about the plaintext images from the encrypted images, while image owners can achieve image usability based on the relevant information about the plaintext images retained in the encrypted images.

Contributors

Xiuhui CHEN processed the data. Yakun MA validated the study. Fang ZUO supervised the study. Xiuli CHAI drafted the paper. Zhihua GAN helped organize the paper. Yushu ZHANG revised and finalized the paper.

Compliance with ethics guidelines

Xiuli CHAI, Xiuhui CHEN, Yakun MA, Fang ZUO, Zhihua GAN, and Yushu ZHANG declare that they have no conflict of interest.

Data availability

The data that support the findings of this study are available from the corresponding authors upon reasonable request.

References

Ashiq JA, 2015. Insider vs. Outsider Threats: Identify and Prevent. INFOSEC. https://resources.infosecinstitute.com/topic/insider-vs-outsider-threats-identify-and-prevent/ [Accessed on Aug. 14, 2022].

Baum LE, Petrie T, 1966. Statistical inference for probabilistic functions of finite state Markov chains. *Ann Math Statist*, 37(6):1554-1563. https://doi.org/10.1214/aoms/1177699147

Beaver D, Kumar S, Li HC, et al., 2010. Finding a needle in haystack: Facebook's photo storage. 9th USENIX Symp on Operating Systems Design and Implementation, p.47-60.

Bellare M, Ristenpart T, Rogaway P, et al., 2009. Format-preserving encryption. Proc 16th Int Workshop on Selected Areas in Cryptography, p.295-312. https://doi.org/10.1007/978-3-642-05445-7_19

Chai XL, Fu JY, Gan ZH, et al., 2022a. An image encryption scheme based on multi-objective optimization and block compressed sensing. *Nonl Dynam*, 108(3):2671-2704. https://doi.org/10.1007/s11071-022-07328-3

Chai XL, Wang YJ, Gan ZH, et al., 2022b. Preserving privacy while revealing thumbnail for content-based encrypted image retrieval in the cloud. *Inform Sci*, 604:115-141. https://doi.org/10.1016/j.ins.2022.05.008

Chen LP, Yin H, Yuan LG, et al., 2021. Double color image encryption based on fractional order discrete improved Henon map and Rubik's cube transform. *Signal Process Image Commun*, 97:116363. https://doi.org/10.1016/j.image.2021.116363

Dong C, Loy CC, Tang XO, 2016. Accelerating the super-resolution convolutional neural network. Proc 14th European Conf on Computer Vision, p.391-407. https://doi.org/10.1007/978-3-319-46475-6_25

Fan LY, 2019. A demonstration of image obfuscation with provable privacy. IEEE Int Conf on Multimedia & Expo Workshops, p.608. https://doi.org/10.1109/ICMEW.2019.00112

Franzese M, Iuliano A, 2019. Hidden Markov models. *Encycl Bioinform Comput Biol*, 1:753-762. https://doi.org/10.1016/B978-0-12-809633-8.20488-3

Hartmann Y, Liu H, Lahrberg S, et al., 2022. Interpretable high-level features for human activity recognition. Proc 15th Int Joint Conf on Biomedical Engineering Systems and Technologies, p.40-49. https://doi.org/10.5220/0010840500003123

He ZL, He YH, Chen LY, 2010. A study on the key issues of cloud storage technology. *Appl Mech Mater*, 29-32:1122-1126. https://doi.org/10.4028/www.scientific.net/AMM.29-32.1122

Jegou H, Douze M, Schmid C, 2008. Hamming embedding and weak geometric consistency for large scale image search. Proc 10th European Conf on Computer Vision, p.304-317. https://doi.org/10.1007/978-3-540-88682-2_24

Jolfaei A, Wu XW, Muthukkumarasamy V, 2016. On the security of permutation-only image encryption schemes. *IEEE Trans Inform Forens Secur*, 11(2):235-246. https://doi.org/10.1109/TIFS.2015.2489178

Joshi AB, Kumar D, Mishra DC, et al., 2020. Colour-image encryption based on 2D discrete wavelet transform and 3D logistic chaotic map. *J Mod Opt*, 67(10):933-949. https://doi.org/10.1080/09500340.2020.1789233

Li SJ, Li CQ, Chen GR, et al., 2008. A general quantitative cryptanalysis of permutation-only multimedia ciphers against plaintext attacks. *Signal Process Image Commun*, 23(3):212-223. https://doi.org/10.1016/j.image.2008.01.003

Marohn B, Wright CV, Feng WC, et al., 2017. Approximate thumbnail preserving encryption. Proc Multimedia Privacy and Security, p.33-43. https://doi.org/10.1145/3137616.3137621

Mishra P, Bhaya C, Pal AK, et al., 2021. A novel binary operator for designing medical and natural image cryptosystems. *Signal Process Image Commun*, 98:116377. https://doi.org/10.1016/j.image.2021.116377

Srivastava RK, Shree R, Shukla AK, et al., 2022. A feature based classification and analysis of hidden Markov model in speech recognition. Proc Cyber Intelligence and Information Retrieval, p.365-379. https://doi.org/10.1007/978-981-16-4284-5_32

Tajik K, Gunasekaran A, Dutta R, et al., 2019. Balancing image privacy and usability with thumbnail-preserving encryption. Network and Distributed Systems Security Symp, p.24-27. https://doi.org/10.14722/ndss.2019.23432

Wang LG, Wang YQ, Dong XY, et al., 2021. Unsupervised degradation representation learning for blind super-resolution. Proc IEEE/CVF Conf on Computer Vision and Pattern Recognition, p.10576-10585. https://doi.org/10.1109/CVPR46437.2021.01044

Wright CV, Feng WC, Liu F, 2015. Thumbnail-preserving encryption for JPEG. Proc 3rd ACM Workshop on Information Hiding and Multimedia Security, p.141-146. https://doi.org/10.1145/2756601.2756618

Wu D, Gan JH, Zhou JX, et al., 2022. Fine-grained semantic ethnic costume high-resolution image colorization with conditional GAN. *Int J Intell Syst*, 37(5):2952-2968. https://doi.org/10.1002/int.22726

Xue TT, Liu H, 2022. Hidden Markov model and its application in human activity recognition and fall detection: a review. Proc 10th Int Conf in Communications Signal Processing and Systems, p.863-869. https://doi.org/10.1007/978-981-19-0390-8_108

Youngblood GM, Cook DJ, 2007. Data mining for hierarchical model creation. *IEEE Trans Syst Man Cybern Part C Appl Rev*, 37(4):561-572. https://doi.org/10.1109/TSMCC.2007.897341

Zhang YS, Zhao RY, Xiao XL, et al., 2022. HF-TPE: high-fidelity thumbnail-preserving encryption. *IEEE Trans Circ Syst Video Technol*, 32(3):947-961. https://doi.org/10.1109/TCSVT.2021.3070348

Zhao RY, Zhang YS, Xiao XL, et al., 2021. TPE2: three-pixel exact thumbnail-preserving image encryption. *Signal Process*, 183:108019. https://doi.org/10.1016/j.sigpro.2021.108019

Zhu Z, Wu C, Wang J, et al., 2020. A novel 3D vector decomposition for color-image encryption. *IEEE Photon J*, 12(2):7800614. https://doi.org/10.1109/JPHOT.2020.2981494