Reversible Data Hiding in JPEG Images under Multi-distortion Metric

Kejiang Chen, Hang Zhou, Dongdong Hou, Weiming Zhang, Nenghai Yu

***Abstract*—Reversible Data Hiding (RDH) in JPEG images is valuable for many applications, such as archive management and image authentication. Recently, there emerged a lot of related works for JPEG RDH, however, the current methods utilize the histogram-shifting-based framework in a constant distortion metric, which does not consider the property of DCT coefficients whose modification distortion is miscellaneous with respect to DCT frequency. In this paper, we propose a novel JPEG RDH scheme under multi-distortion metric. At first, the modification distortion is defined as the impact in the spatial domain caused by modifying DCT coefficients. Different from previous works, here we select coefficients with values ±1, ±2 as cover, which will not cause extra bitrate-expansion and decrease the shift distortion. By minimizing both the modification distortion and shift distortion, a heuristic block selecting strategy is proposed. With the selected coefficients and the corresponding distortion, we use recursive histogram modification under the inconsistent-distortion metric for message embedding. The experimental results show that the proposed method can effectively improve visual quality as well as bring small bitstream expansion. Additionally, the undetectability of the proposed method outperforms those of the existing methods.**

***Index Terms*—reversible data hiding, JPEG, multi-distortion**

I. Introduction

REVERSIBLE Data Hiding (RDH) embeds message into the host signal with the aim of minimizing the distortion incurred and can losslessly recover the host signal from the marked image [1]. It is widely applied in medical images, military images and law forensics, where the host signal is required not to be damaged.

In the past decades, various RDH schemes have been proposed that can be roughly classified into three major strategies: lossless compression appending [2], [3], Difference Expansion (DE) [4]–[6] and Histogram Shifting (HS) [7], [8]. Most of the RDH methods are designed for spatial images, and many considerable achievements have been made. When it comes to the compressed domain, there are only a few works. Actually, the essence of RDH is compression, which compresses the host signal to leave room for secret messages. Compared with spatial images, the compressed images have less redundancy, thus it is considerably more difficult to design RDH methods for compressed images [9]. However, as the most popular format used in digital cameras and other photography capturing devices, JPEG images are enormously used in medical and military fields, which show great potential value in archive management and image authentication.

Recently, RDH in JPEG images has received increasing attention and three major approaches have been developed for the RDH into JPEG images: 1) coefficient-based RDH [10]–[15], 2) quantization-table-based RDH [2], [16]–[19], 3) Huffman-codes-based RDH [20]–[24]. Since the capacities of the latter two categories are rather limited, we focus on the RDH based on the manipulation of quantized DCT coefficients.

Fridrich *et al.* [2] proposed the idea of losslessly compressing the Least Significant Bit (LSB) of the selected DCT coefficients in a JPEG image to vacate space for RDH. Xuan *et al.* [10] proposed an RDH method for JPEG images by using histogram pairs. This method shifts the quantized DCT coefficient histogram with an optimal search strategy to minimize the modification distortion. Efimushkina *et al.* [13] embedded messages into selected coefficients with small magnitudes to ensure that little distortion is introduced by modification in the host JPEG image. Nikolaidis [14] proposed an RDH scheme for JPEG image by relying on the modification of the quantized DCT coefficients with the value of zero.

Huang *et al.* [15] proposed an HS-based RDH scheme for JPEG image, by which only Alternating Current (AC) coefficients with values "±1" are selected as host to carry messages. Moreover, a strategy for block selection is proposed based on the number of zero AC coefficients in each 8×8 block, so that the method can achieve high embedding capacity and preserve the visual quality and the file size of the marked JPEG image. Utilizing the number of zero coefficients to describe the distortion introduced by embedding is not precise enough. Hou *et al.* [25] improved the selection strategy by selecting DCT block according to the unit distortion introduced by simulating the embedding process. From another perspective, Hong *et al.* [26] improved the performance by embedding the message into a continuous interval with large correlation and low distortion.

In previous methods, after selecting blocks, the DCT coefficients are treated equally. That is to say, the modification impacts caused by different coefficients are all the same. Obviously, this is not the optimal choice. Modifying coefficients in low frequencies causes smaller perturbations in the spatial domain than that in high frequencies.

To solve this, we propose an RDH method for JPEG images under inconsistent distortion metric in this paper. At first, the

This work was supported in part by the Natural Science Foundation of China under Grant U1636201, by the Youth Program of National Natural Science Foundation of China under Grant 62002334, by the Anhui Science Foundation of China under Grant 2008085QF296, and by the Anhui Initiative in Quantum Information Technologies under Grant AHY150400.

All the authors are with CAS Key Laboratory of Electro-magnetic Space Information, University of Science and Technology of China, Hefei 230026, China.

Corresponding author: Weiming Zhang (Email: zhangwm@ustc.edu.cn)

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2020.3044186, IEEE Transactions on Circuits and Systems for Video Technology

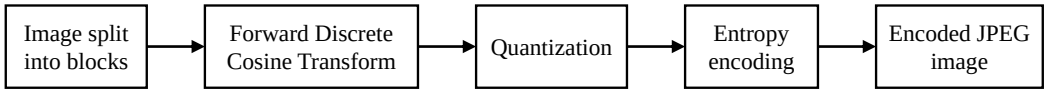


Figure 1. The baseline sequential JPEG encoding process.

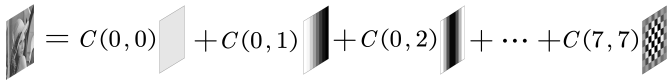


Figure 2. An example of a linear combination of DCT basis images.

modification distortion of cover coefficients is defined, which is related to the quantization step. Then suitable coefficients are selected as host to pursue a tradeoff between distortion of the inner region and the outer region. Given the selected coefficients, the corresponding distortion and message, the optimal transition probability matrix (OTPM) can be estimated using the method proposed in [27]. With OTPM, the RDH can be well implemented by Recursive Histogram Modification under multi-distortion metric (M-RHM) [28].

The main contributions of this paper are listed as follows:

1) Considering modifying different coefficients will cause different impacts, we define reproductive distortion functions based on spatial pixel difference according to the linearity of the DCT. To the best of our knowledge, this is the first work treating DCT coefficients differently when embedding the message in the coefficient-based JPEG RDH.
2) Different from previous works, in addition to ±1 we also select coefficients with the value of ±2 as cover, which will not cause more bit expansion as well as decrease the shift distortion. Moreover, a strategy of selecting suitable blocks is also proposed, which pursues a tradeoff between the distortion of the inner region and the outer region.
3) The proposed method achieves better visual quality, less bit-rate expansion and higher undetectability compared to the existing methods.

The rest of this paper is organized as follows. In Section II, we review the process of JPEG compression and related works. The proposed method is elaborated in Section III. The results of comparative experiments are presented in Section IV. Conclusion and future work are given in Section V.

II. Related Works

A. Overview of JPEG Compression

JPEG is a commonly used method of lossy compression for digital images, particularly for those images produced by digital photography. The baseline sequential JPEG encoding process of grayscale images is presented in Fig. 1, which is performed in three sequential steps: DCT computation, quantization, and variable-length code assignment [29].

DCT expresses a finite sequence of data points in terms of a sum of cosine functions oscillating at different frequencies. As shown in Fig. 2, the spatial image block $\mathbf{X}$ is decomposed as a linear combination of DCT basis images of different frequencies weighted by the DCT coefficients:

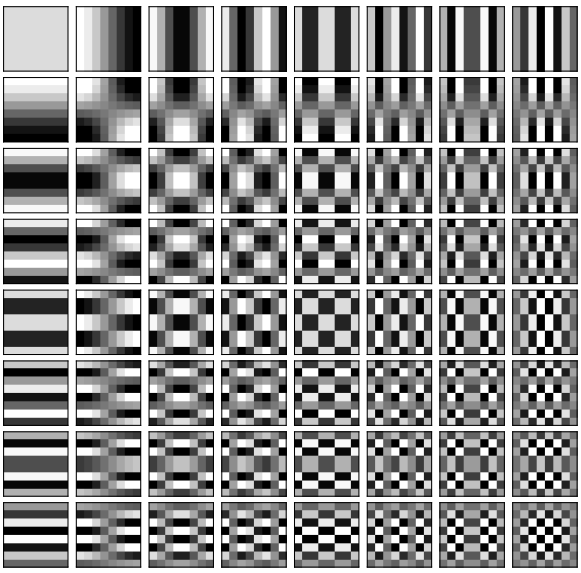


Figure 3. The DCT basis images. Each step from left to right and top to bottom is an increase in frequency by 1/2 cycle.

$$\mathbf{X} = \sum_{k,l} C(k,l)\mathbf{B}_{k,l}. \tag{1}$$

$C(k,l)$ is the unquantized DCT coefficients and $\mathbf{B}_{k,l}$ is the DCT basis image:

$$\mathbf{B}_{k,l} = \begin{bmatrix} \mathbf{f}(0,0;k,l) & \cdots & \mathbf{f}(0,7;k,l) \\ \vdots & \ddots & \vdots \\ \mathbf{f}(7,0;k,l) & \cdots & \mathbf{f}(7,7;k,l) \end{bmatrix} \tag{2}$$

where

$$\mathbf{f}(i,j;k,l) = \frac{\mathbf{w}[k]\,\mathbf{w}[l]}{4} \cos\frac{\pi}{16}k(2i+1)\cos\frac{\pi}{16}l(2j+1), \tag{3}$$

and $\mathbf{w}[0] = 1/\sqrt{2}, \mathbf{w}[k>0] = 1$. Fig. 3 gives the visualization of DCT basis images.

To reduce the amount of information, the compressor will divide each DCT coefficient value by a quantization step and round the result to an integer. The quantization step controls the compression ratio, and larger values lead to greater compression.

$$\hat{C}(k,l) = \text{round}\left(\frac{C(k,l)}{Q(k,l)}\right). \tag{4}$$

The Independent JPEG Group (IJG) specifies a standard quantization table for a quality factor $QF = 50$, as shown

Table I
THE QUANTIZATION TABLE **Q** OF QF = 50

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	57	64	81	104	113	92
49	64	78	87	103	121	120	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

in Table I. From this table, each quantization table with a quality factor $QF \in [1, 100]$ can be calculated [30].

$$\mathbf{Q}_{QF} = \begin{cases} \max\{\mathbf{1}, \text{round}\,(2\mathbf{Q}_{50}\,(1 - QF/100))\}, & QF > 50 \\ \min\{255 \cdot \mathbf{1}, \text{round}\,(\mathbf{Q}_{50}50/QF)\}, & QF \leqslant 50 \end{cases} \tag{5}$$

After that, the entropy coding will be implemented to obtain the JPEG bitstream.

B. Histogram-Shifting(HS)-based JPEG RDH

The histogram-shifting-based RDH is one of the most popular RDH schemes, benefiting from its high visual quality and large capacity. It usually consists of three steps: generating the histogram from the host signal, splitting the histogram into an inner region and an outer region, and embedding messages by modifying the inner bins as well as shifting the outer bins. Analogous to the relation between the entropy of source data and the theoretical limit to data compression, the entropy of histogram determines the limit to embedding capacity [31]. As a result, mainstream schemes are trying to produce sharp histogram owning small entropy.

Huang *et al.* [15] proposed an HS-based RDH scheme for JPEG images that embeds message into quantized AC coefficients, whose histogram is sharp. Without any loss of generality, the non-zero AC coefficients are represented as $\mathbf{c} = \{c_1, c_2, \ldots, c_N\}$, where N is the total number of the non-zero AC coefficients in the JPEG image. The process of message embedding is denoted as follows:

$$\tilde{c}_i = \begin{cases} c_i + \text{sign}\,(c_i) * b & \text{if } |c_i| = 1 \\ c_i + \text{sign}\,(c_i) & \text{if } |c_i| > 1 \end{cases} \tag{6}$$

where

$$\text{sign}(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -1 & \text{if } x < 0 \end{cases} \tag{7}$$

In Eq. (6), b denotes the message bit to be embedded, and $\tilde{c}_i$ represents the marked signal after embedding. The message extraction and host signal recovery process are as follows:

$$b' = \begin{cases} 0 & \text{if } |\tilde{c}_i| = 1 \\ 1 & \text{if } |\tilde{c}_i| = 2 \end{cases} \tag{8}$$

$$c'_i = \begin{cases} \text{sign}\,(\tilde{c}_i) & \text{if } 1 \leq |\tilde{c}_i| \leq 2 \\ \tilde{c}_i - \text{sign}\,(\tilde{c}_i) & \text{if } |\tilde{c}_i| \geq 3 \end{cases} \tag{9}$$

where b' and c'_i denote the extracted message bit and the recovered signal, respectively. Specifically, the AC coefficients that satisfy $|c_i| = 1$ belong to the inner region, and those satisfy $|c_i| > 1$ comprise the outer region.

Since too many blocks will introduce unnecessary distortion from the outer region, selecting proper coefficients counts for a lot. The strategy of the coefficient selection is always divided into two parts: selecting frequencies and selecting blocks. In Huang *et al.*'s method [15], the nonzero AC coefficients for all frequencies are selected for data hiding and blocks with more zero coefficients will take precedence in the embedding process. In detail, all the blocks in which the number of zero AC coefficients is not less than the threshold value T_z are selected. The threshold T_z is determined as:

$$T_z = \underset{T}{\text{argmax}}\,\{S \geq (L + l_1 + l_2)\}, \tag{10}$$

where $0 \leq T \leq 63$, L denotes the number of message bits to be embedded, l_1 represents the number of bits needed to represent the value of L, and l_2 indicates the number of bits needed to represent the threshold value T_z.

Hou *et al.* [9] selected coefficients from frequencies yielding less distortion for embedding, and then an advanced block selection strategy is applied to always modify the block yielding less simulated distortion firstly until the given payloads are completely embedded. Hong *et al.* [26] proposed a novel strategy of selecting DCT coefficients based on the correlation among images.

C. RHM under Multi-distortion Metric

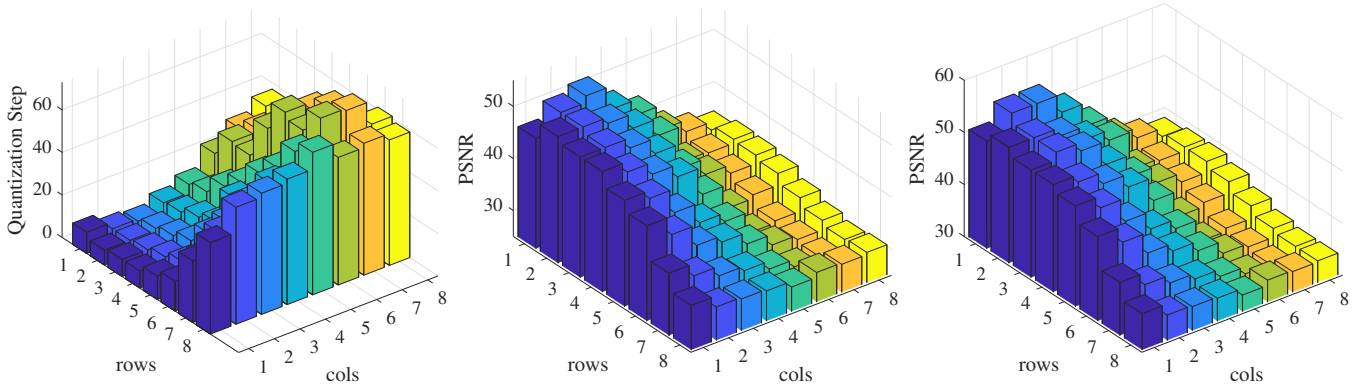
In our proposed scheme, we will utilize a minimal distortion method to embed the message. Zhang *et al.* [27] proposed the optimization problem that deals with minimal distortion problem under Payload-Limited Sender (PLS) for RDH schemes.

$$\begin{array}{ll} \text{minimize} & \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} P_X(x) P_{Y|X}(y|x) d(x, y) \\ \text{subject to} & H(Y) = R + H(X) \end{array} \tag{11}$$

where X and Y denote the random variables of the host sequence and the marked sequence, $P_{Y|X}(y|x)$ is the transition probability, $H(\cdot)$ is the information entropy function. Subsequently, they proposed a general framework for estimating the optimal transition probability matrix (OTPM) $P_{Y|X}(y|x)$ from $P_X(x)$ and $P_Y(y)$ under any consistent distortion metrics [27]. According to the OTPM, message can be reversibly embedded and the average distortion can be minimized with an RCC (recursive code construction) scheme such as RHM (recursive histogram modification) [28].

The single distortion metric work is extended by Hou *et al.* and they designed method for estimating OTPM for RDH under a multi-distortion metric [25]. At first, the elements' distortion is clustered into K classes. Correspondingly, the host sequence is divided into K subsequences, denoted by $\mathbf{X}_i = (x_{i,1}, x_{i,2}, \cdots, x_{i,N_i})$, where N_i is the length of $\mathbf{X}_i$. All elements in the subsequence $\mathbf{X}_i$ will share the same distortion metric defined as $d_i(x, y)$. For each subsequence $\mathbf{X}_i$, the embedding rate is denoted by R_i (relative to $\mathbf{X}_i$), then the corresponding sub-OTPM denoted as $P_{Y_i|X_i}(y|x)$ can be obtained using the previous work (OTPM for the consistent distortion metric), and then modify the histogram of the host signal to obtain the corresponding marked signal, denoted by $\mathbf{Y}_i$. The total

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2020.3044186, IEEE Transactions on Circuits and Systems for Video Technology



(a) Quantization Table (b) QF = 70 (c) QF = 80

Figure 4. The PSNR values corresponding to changing coefficients of different DCT frequencies. (a) Quantization Table. (b) QF = 70. (c) QF = 80.

distortion caused by embedding the payload R_i into $\mathbf{X}_i$ (denoted by J_i) is

$$J_i = N_i \sum_{x,y} P_{X_i}(x) P_{Y_i|X_i}(y|x) d_i(x,y). \tag{12}$$

As for entire payload R, the corresponding payload carried by the subsequence $\mathbf{X}_i$ is $\frac{N_i \times R_i}{N}$. As a result, the sum payload of the whole sequence equals to

$$R = \frac{\sum_{i=1}^{K} N_i \times R_i}{N}. \tag{13}$$

The optimizing problem of multi-distortion OTPM for RDH under PLS can be formulated as:

$$\begin{aligned} &\text{minimize} \quad \frac{\sum_{i=1}^{K} N_i \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} P_{X_i}(x) P_{Y_i|X_i}(y|x) d_i(x,y)}{N} \\ &\text{subject to} \quad \frac{\sum_{i=1}^{K} N_i \times H(Y_i)}{N} = R + \frac{\sum_{i=1}^{K} N_i \times H(X_i)}{N} \end{aligned} \tag{14}$$

A similar solution can be put forward for this problem as the single distortion by assigning infinite distortion to inter-class modification patterns.

III. Proposed Method

A. Motivation

The existing RDH schemes based on the manipulation of quantized DCT coefficients considering the priorities of modifying coefficients are the same, which is obviously unreasonable, because the impacts of modifying coefficients in different frequencies are different. Due to the orthogonality and linearity of the Discrete Cosine Transform, the impact of modifying different coefficients is additive. Therefore, the impact through modifying a coefficient on the cover image is equivalent to the impact of modifying a coefficient on the all-zero DCT block.

We calculate the peak signal-to-noise ratio (PSNR) between the all-zero DCT block and the modified block. The modified block is obtained by changing kl-th coefficients ($k = 0, 1, ..., 7, l = 0, 1..., 7$) from zero to one on the all-zero DCT block. Fig. 4 (a) is the quantization table, and Fig. 4 (b) and (c) present PSNR values of changing coefficients in different frequencies with QF = 70 and 80, respectively. It can be observed that modification on the coefficients with low frequency owns higher PSNR.

Here, we also give theoretical analysis of this conclusion. PSNR is defined as follows:

$$PSNR = 10 \cdot \log_{10}\left(\frac{MAX_I^2}{MSE}\right), \tag{15}$$

where MSE is the mean square error of spatial images:

$$MSE = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \|I'(i,j) - I(i,j)\|^2, \tag{16}$$

where I and I' are the original image and modified image, respectively. M, N are the height and width of the image. According to Parseval's Theorem, the MSE in the pixel domain is equivalent to the DCT domain mean square quantization error (MSQE), because DCT is a normalized orthogonal transform [32]. Therefore, the PSNR can also be expressed as:

$$PSNR = 10 \cdot \log_{10}\left(\frac{MAX_I^2}{MSQE}\right), \tag{17}$$

where

$$MSQE = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \|F'(i,j) - F(i,j)\|^2, \tag{18}$$

where F and F' are the unquantized DCT coefficients of the original image I and the modified image I'. Due to the linearity of the DCT, we can measure the impact of the spatial domain by calculating PSNR between the all-zero block and its modified version. For example, if we modify (k, l) coefficient, the $MSQE$ becomes:

$$MSQE = \frac{1}{MN} Q(k,l)^2. \tag{19}$$

Then we have

$$PSNR = 10 \cdot \log_{10}\left(\frac{MAX_I^2 MN}{Q(k,l)^2}\right). \tag{20}$$

It can be seen that modifying coefficients with low quantization step own higher PSNR. Consequently, we tend to treat different coefficients under multi-distortion metric.

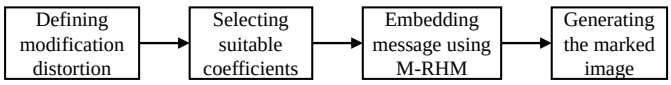
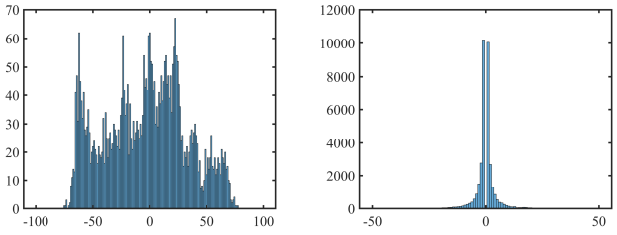


Figure 5. The pipeline of the proposed method.



(a) DC coefficients (b) Non-zero AC coefficients

Figure 6. Histograms of all DC coefficients (a) and non-zero AC coefficients (b) of *Lena* image, QF=70.

B. Overview of the Proposed Method

In this section, we elaborate a novel scheme of JPEG reversible data hiding under multi-distortion metric. We overview the framework of the proposed method firstly. The pipeline of the proposed scheme is presented in Fig. 5. The method consists of four modules: defining modification distortion, selecting suitable coefficients, embedding message, and generating the final marked image. At first, the modification distortion is defined for DCT coefficients which is related to the quantization step. Then, the quantized AC coefficients with value of ±1,±2 are picked as the cover sequence, and coefficients with the amplitude larger than 2 are shifted for creating a sharp histogram, which constitutes the outer region. Thereafter, the average payload of the coefficients and the needed blocks are determined by minimizing the distortion of the inner region and the outer distortion. Finally, the message embedding, message extraction, and host recovery will be implemented by recursive histogram modification under multi-distortion metric (M-RHM) [25]. Details of the modules will be introduced below.

C. Distortion Definition

Visual quality is an important factor on measuring the performance of RDH schemes. Most of subjective and objective visual quality metrics are evaluated in the spatial domain. Therefore, we design distortion based on the spatial impacts caused by the modification on DCT coefficients. It is worth noting that M-RHM is adopted for message embedding, which requires that the distortion calculated by the receiver should be the same as that calculated by the sender. This property means the distortion should be independent of the image content.

According the linearity of the DCT, the impacts in the spatial domain of modifying different coefficients are *additive*. Below gives the corresponding proof extended from Eq. (1). The difference in the spatial domain between the modified JPEG image block $\mathbf{X}'_{k'l'}$ ($k'l'$-th coefficient modified) and the original JPEG image block $\mathbf{X}$ is

$$\begin{aligned}\mathbf{D}_{k'l'} &= \mathbf{X}'_{k'l'} - \mathbf{X} \\ &= \sum_{k,l}\Big(\hat{C}(k,l) + \delta(k-k', l-l')\Big)Q(k,l)\mathbf{B}_{k,l} \\ &\quad - \sum_{k,l}\hat{C}(k,l)Q(k,l)\mathbf{B}_{k,l} \\ &= \sum_{k,l}\delta(k-k', l-l')Q(k,l)\mathbf{B}_{k,l} \\ &= Q(k',l')\mathbf{B}_{k',l'}\end{aligned} \tag{21}$$

where δ is the Kronecker delta. It is clear that the difference in the spatial domain is not related to the image content but the quantization step and the DCT basis image.

With this prior knowledge, we design a concise distortion as the MSE in the spatial domain caused by changing DCT coefficients. Concretely, the distortion D_{kl} of modifying kl-th coefficient in unit 1 on the cover image can be measured by the square of ℓ_2-norm (MSE) of the spatial difference:

$$D_{kl} = \left\|Q(k,l)\mathbf{B}_{k,l}\right\|_2^2. \tag{22}$$

It can be seen that coefficients with different quantization steps own different distortion and the larger the quantification step, the larger the distortion, which corresponds to our motivation.

As for modifying coefficient in larger amplitude α, such as ±2 and ±3, the distortion D_{kl}^{α} is defined as:

$$D_{kl}^{\alpha} = \alpha^2 \cdot D_{kl}. \tag{23}$$

To improve the undetectability, we design a precise distortion function based on the wavelet filter bank residual (named WFB). The underlying reason is that the strong JPEG steganalysis methods (DCTR [33], GFR [34]) are constructed on the filter residuals. *The larger the filter residual change caused by DCT coefficients modification, the larger the distortion.*

In implementation, the filter bank $\mathcal{B}_n = \left\{\mathbf{K}^{(1)}, ..., \mathbf{K}^{(n)}\right\}$ consists of multiple directional wavelet high-pass filters represented by their kernels normalized so that all ℓ_2-norms $\|\mathbf{K}^{(i)}\|_2$ are the same. The filter residual of changing kl-th coefficient in unit 1 is:

$$\mathbf{R}_{kl}^{(i)} = \mathbf{K}^{(i)} \star Q(k,l)\mathbf{B}_{k,l}, \tag{24}$$

where $\star$ is mirror-padded convolution. Then the distortion of changing the kl-th coefficient in unit 1 is defined as:

$$D_{kl} = \left(\sum\nolimits_i \left\|\mathbf{R}_{kl}^{(i)}\right\|_2\right)^{\frac{1}{2}}. \tag{25}$$

It can be seen that the distortion is positively correlated with the filter residual. The purpose of the root square is to avoid too large differences of the distortion. Otherwise, the probability differs widely and the information entropy of the probability is low, resulting in large modifications. Daubechies 8 (db8) wavelet filter bank [35] is adopted as the filter bank, which consists of three filter kernels:

$$\mathbf{K}^{(1)} = \mathbf{h}\cdot\mathbf{g}^{\mathrm{T}}, \mathbf{K}^{(2)} = \mathbf{g}\cdot\mathbf{h}^{\mathrm{T}}, \mathbf{K}^{(3)} = \mathbf{g}\cdot\mathbf{g}^{\mathrm{T}}, \tag{26}$$

where $\mathbf{h}$ is the decomposition low-pass filter and $\mathbf{g}$ is the decomposition high-pass filter, as shown in Fig. 7.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2020.3044186, IEEE Transactions on Circuits and Systems for Video Technology

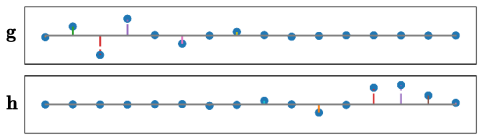


Figure 7. The Daubechies 8 wavelet filters (top: decomposition high-pass filter **g**; bottom: decomposition low-pass filter **h**).

D. Selecting Suitable Coefficients

1) Selecting Suitable Amplitude: The histogram of candidate coefficients ought to be sharp in the HS-based RDH schemes. We calculated the histogram of all quantized Direct Coefficient (DC) coefficients and non-zero AC coefficients. Zero AC coefficients are not in the statistical range, because the modification of zero coefficients will result in a disproportion increase in the JPEG file size [15]. The histogram of all quantized DC coefficients is rather flat and that of quantized non-zero AC coefficients is sharp, as shown in Fig. 6. Consequently, quantized non-zero AC coefficients are selected as the candidate host. Furthermore, the height of the coefficient bin decides the message capacity, and the amount of the coefficients with low amplitude always turns to be large. From the above analysis, the coefficients with low amplitude will be selected as the host signal. Different from the previous work in which the quantized AC coefficient with the values of ± 1 are chosen, the quantized AC coefficient with the values of $\pm 1, \pm 2$ will be picked as the host signal for message embedding using M-RHM in the proposed method. Then the coefficients with the amplitude larger than 2 comprise the outer region, as formulated in Eq. (27).

$$\tilde{c}_i = \begin{cases} \text{M-RDH}(c_i) & \text{if } 1 \leqslant |c_i| \leqslant 2 \\ c_i + \text{sign}(c_i) & \text{if } |c_i| > 2 \end{cases} \tag{27}$$

The underlying reasons of selecting the quantized AC coefficient with the values of $\pm 1, \pm 2$ as cover sequence are listed as follows:

- **No extra bitstream expansion:** After quantization, run length encoding is carried out for quantized AC coefficients in JPEG compression. The run-length coder generates a codeword ((*run-length*, *category*), *amplitude*), where *run-length* is the length of zero run followed by the given non-zero coefficient, *amplitude* is the value of this non-zero coefficient and *category* is the number of bits needed to represent the *amplitude*, which is shown in Table II. As shown, the coefficients with the values of $\pm 2, \pm 3$ are in the same *category*, meaning that the transfer from ± 2 to ± 3 will not cause extra bitstream expansion.
- **Less distortion of the outer region:** Selecting more kinds of amplitudes means more coefficients in one block can be used as the cover. Naturally, the payload of one block will increase, such that fewer blocks need to be selected as the host sequence. Therefore, less distortion of the outer region will be introduced.

2) Selecting Suitable Blocks: After choosing the amplitude of the host, the subsequent process is determining the payload and the blocks we need for the given message. It is self-evident that a moderate amount of blocks do good to visual quality.

Table II
CATEGORIES ASSIGNED TO COEFFICIENT VALUES

C	AC Coefficients
0	0
1	-1,1
2	-3,-2,2,3
3	-7,...,-4,4,...7
4	-15,...,-8,8,...15
5	-31,...-16,16,...31
6	-63,...-32,32,...63
7	-127,...-64,64,...127
8	-255,...-128,128,...255
9	-511,...-256,256,...511
10	-1023,...-512,512,...1023

In the traditional HS-based RDH scheme, one coefficient can carry 1-bit message. If we use M-RHM for embedding, the maximum capacity of one coefficient can reach $\log_2 N_c$, where N_c represents how many values are adopted as the host signal. The maximum capacity means that the modification possibilities are the same, which violates our original intention: coefficients of different DCT frequencies own different modifying priorities. Inspired by the content-adaptive steganography [36]–[38], the secure capacity of one element is always quite smaller than the maximum capacity, which keeps the modification distortion in a low level. However, a smaller payload means more blocks will be selected as the cover, which will increase the distortion of the outer region. Consequently, we should find a tradeoff between the distortion of the inner region and the outer region. This problem can be formulated as the following optimization problem:

$$\begin{aligned} &\text{minimize} \quad \rho_{\text{in}}(R, X) + \rho_{\text{out}}(R, X) \\ &\text{subject to} \quad RX = N \end{aligned} \tag{28}$$

where $\rho_{\text{in}}, \rho_{\text{out}}$ are the distortion of the inner region and the outer region, respectively. R is the average payload, X is the cover sequence composed of the coefficients with values $\pm 1, \pm 2$ in the selected blocks $\mathcal{B}$, and N is the length of the message. In practice, however, the constraint of the reversibility and the discrete value of X prevent us from finding an analytical solution of Eq. (28). As an alternative, we set a moderate payload and select blocks heuristically. In the implementation, we obtain the optimal R by traversal search. R ranges from 0.1 to N_c, and is determined by comparing PSNR between the original images and the marked images. Given the average payload R, the detailed steps for collecting coefficients are listed as follows:

Step 1: Given a JPEG file, we first entropy-decode it and obtain quantized DCT coefficients. Thereafter, compute the ρ_{out} for each 8×8 block and sort blocks in ascending order with respect to ρ_{out}, denoted by $B_1, B_2, ..., B_{N/64}$. To point out, the ρ_{out} of every block is independent of the message, because the modification pattern is fixed. As a result, the receiver can produce the same result as the sender.

Step 2: Compute the number of the quantized AC coefficients with the values of ± 1 and ± 2 in each block following the mentioned order, denoted by $L_1, L_2, ...L_{N/64}$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2020.3044186, IEEE Transactions on Circuits and Systems for Video Technology

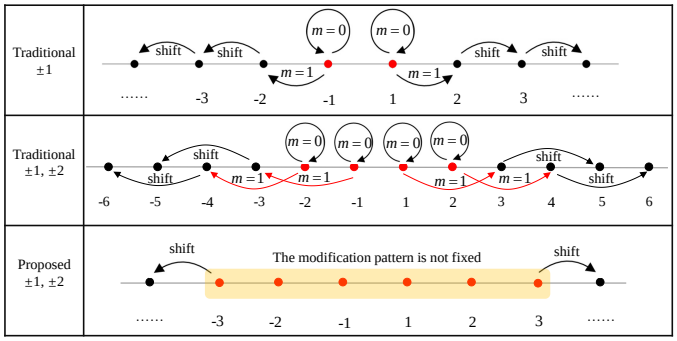


Figure 8. The modification patterns of previous RDH schemes. The traditional $\pm 1, \pm 2$ modification pattern needs more shift distortion (Row 2) than that of the proposed scheme using M-RHM (Row 3).

Step 3: Increase the number of blocks p, until satisfying payload restraint:

$$\sum_{i=1}^{p} RL_i \geq M, \tag{29}$$

where M is the length of the message.

Step 4: Collect the coefficients whose values are $\pm 1, \pm 2$ in the selected blocks $\mathcal{B} = \{B_1, B_2, ..., B_p\}$ as the final host.

E. Message Embedding and Extraction using M-RHM

1) Message Embedding: With the selected coefficients and the corresponding distortion, the next process is embedding the message while minimizing the distortion. M-RHM provides a methodology under the recursive framework that embeds the message with the performance approaching the rate-distortion bound as well as ensures reversibility. Besides, using M-RHM for message embedding will cause less shift distortion than using the traditional histogram shifting method. The intuitive explanation of the inference is given Fig. 8.

The embedding process is listed as follows:

Step 1: Cluster the selected coefficients $\mathbf{X}$ as well as the corresponding distortion $\mathbf{D}$ into K classes according to the distortion $\mathbf{D}$ using k-Nearest Neighbours (k-NN), and obtain the subsequence $\mathbf{x}_i$ and its corresponding distortion $\mathbf{D}_i$ for $i = 1, 2, ..., K$.

Step 2: Concatenate the subsequence $\mathbf{x}_i$ into the compound signal $\mathbf{X}_c$, and then normalize the corresponding histogram $\mathbf{H}_c$ to obtain the compound host probability distribution $P_{\mathbf{X}_c}$.

Step 3: Construct the transfer distortion matrix $\mathbf{D}_c$, which forbids inter-class modification by assigning infinite distortion to these modification patterns:

$$\mathbf{D}_c = \begin{bmatrix} \mathbf{D}_1 & \infty & ... & \infty \\ \infty & \mathbf{D}_2 & ... & \infty \\ \vdots & \vdots & \ddots & \vdots \\ \infty & \infty & ... & \mathbf{D}_K \end{bmatrix} \tag{30}$$

Step 4: With $\mathbf{P}_{X_c}$, $\mathbf{D}_c$ and the embedding rate R as parameters, estimate the OTPM $\mathbf{P}_{Y_c|X_c}$ by solving the optimization problem in Eq. (14) using method in [27].

Step 5: Embed the message $\mathbf{m}$ with OTPM $\mathbf{P}_{Y_c|X_c}$ by recursively applying the decoding process of an entropy coder (RHM) [28], which can be simplified as:

$$\mathbf{Y}_c = Dec(\mathbf{m}, \mathbf{P}_{Y_c|X_c}, \mathbf{X}_c), \tag{31}$$

where $Dec()$ represents the recursive decoder of an entropy coder, and we use arithmetic coding in implementation. For details, the readers can refer to [28].

Step 6: Divide the $\mathbf{Y}_c$ into the subsequence $\mathbf{y}_i$ and then combine $\mathbf{y}_i$ to $\mathbf{Y}$ in the inverse process of Step 1, 2.

For ensuring reversibility, auxiliary information, such as the compound host probability distribution $\mathbf{P}_{X_c}$, the embedding rate R and the number of classes K, is necessary for constructing OTPM, so we should vacate some elements as the reserve region for embedding the auxiliary message using the Least Significant Bit (LSB) substitution. The substituted bits will be a new part of the message. In the proposed scheme, if the length of raw messages to be M bits, we will allocate $0.03M$ bits for auxiliary messages. Then the rest elements compose the host signal $\mathbf{X}$, and the payload is $R = \frac{1.03M}{N_h}$, where N_h is the number of $\mathbf{X}$. Entropy encode the substituted coefficients to obtain the marked JPEG file J_s.

2) Message Extraction and Host Signal Recovery: The receiver fist entropy-decode the marked file J_s, extract the auxiliary information. After that, selecting the DCT coefficients according to the strategy used in the embedding process. The remaining steps of message extraction and host signal recovery are listed as follows:

Step 1: Extract auxiliary information from the reserved elements, including the compound host probability distribution $\mathbf{P}_{X_c}$, the embedding rate R, and the number of class K.

Step 2: Calculate the distortion $\mathbf{D}$ of coefficients based on Eq. (23).

Step 3: Divide the marked sequence $\mathbf{Y}$ and the corresponding distortion $\mathbf{D}$ into K classes according to the distortion $\mathbf{D}$, and obtain the marked subsequence $\mathbf{y}_i$ and the corresponding distortion $\mathbf{D}_i$, for $i = 1, 2, ..., K$.

Step 4: Concatenate the subsequence $\mathbf{y}_i$ for $i = 1, 2, ..., K$ together to generate the compound marked sequence $\mathbf{Y}_c$.

Step 5: Construct the transfer distortion matrix $\mathbf{D}_c$ based on Eq. (30).

Step 6: Using $\mathbf{P}_{X_c}$, $\mathbf{D}_c$ and embedding rate R as parameters, calculate the OTPM $\mathbf{P}_{Y_c|X_c}$.

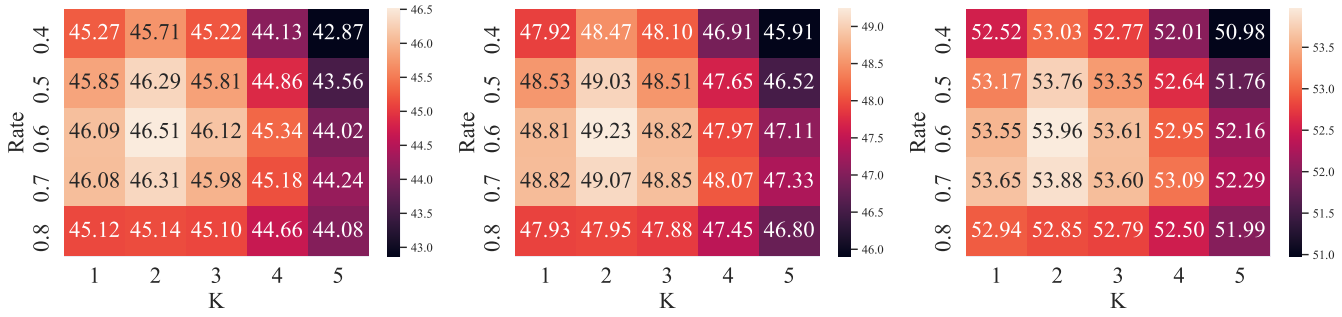
Step 7: The message $\mathbf{m}$ and the compound signal $\mathbf{X}_c$ can be recovered using the encoder $Enc()$ of the entropy coder in a recursive way:

$$\mathbf{m}, \mathbf{X}_c = Enc(\mathbf{Y}_c, \mathbf{P}_{Y_c|X_c}, R). \tag{32}$$

Step 8: Reassemble the compound signal $\mathbf{X}_c$ to the original selected coefficients $\mathbf{X}$ in the inverse process of Step 2 and Step 3.

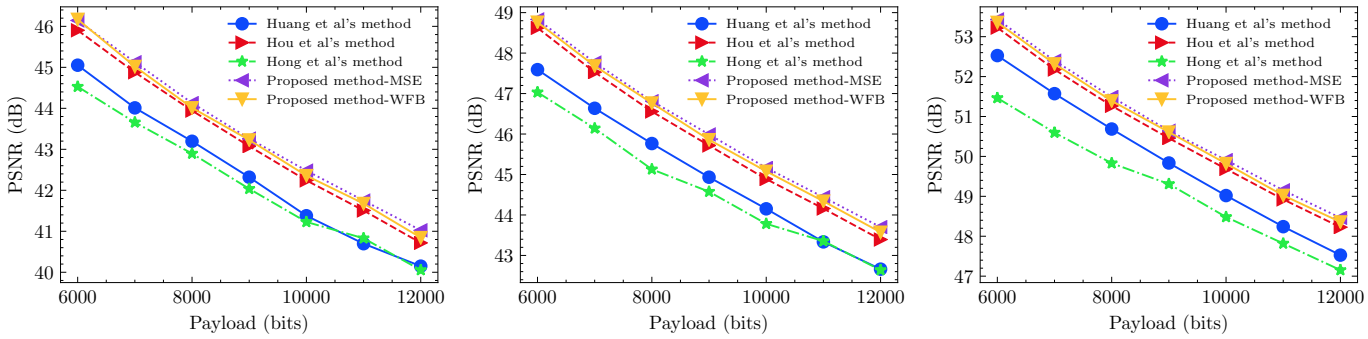
Extract the LSB of the reserved region from the message $\mathbf{m}$, and recover them to the substitute elements. Then entropy encode the selected coefficients $\mathbf{X}$ and the unselected coefficients to obtain the original JPEG file.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2020.3044186, IEEE Transactions on Circuits and Systems for Video Technology



(a) QF=70 (b) QF=80 (c) QF=90

Figure 9. The PSNR values with respect to different K and R. $K = 2$ and $R = 0.6$ outperform others with respect to different quality factors: 70, 80 and 90.



(a) QF = 70 (b) QF = 80 (c) QF = 90

Figure 10. The average **PSNR** values corresponding to different embedding payloads.

IV. Experiments

In all our experiments, the Alaska Dataset [39] and Bossbase 1.01 [40] are adopted as the testing database. The secret message bits are randomly generated and the 512×512 JPEG images are compressed with MATLAB function *imwrite*. The randomly selected 100 images in Bossbase are used for parameters setting, and another disjoint 100 images from Bossbase 1.01 are adopted for performance evaluation. Since different methods will cause different failures, the performance is evaluated on the intersection of successful samples of all methods. 10,000 selected images from Alaska Dataset is used for Steganalytic experiments. To validate the effectiveness of the proposed method, three state-of-the-art RDH schemes [9], [15], [26] for JPEG images are selected for comparison.

Two aspects, namely visual quality and file size preservation, are discussed. The peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) [41], calculated between the original JPEG image and the marked JPEG image, are used as measures to evaluate the visual quality of the marked JPEG image. The increased file size is used to measure the ability of file size preservation.

A. Determining Parameters

There are two parameters that should be be determined, one is R, the average payload, the other one is K, the number of the classes of distortion. The empirical average payload R is obtained through grid search, and the performance is evaluated by PSNR on the MSE distortion.

When the payload R is too small or too large, the embedding of the proposed method will fail. When the R is too small, the message restraint in Eq. (29) cannot be achieved. When the R is too large, namely, $R > \max(H(Y)) - H(X)$, the OTPM cannot be obtained for not satisfying the constraint in the optimization problem in Eq. (11). Here, we explore the performance of payload R ranging 0.4 to 0.8.

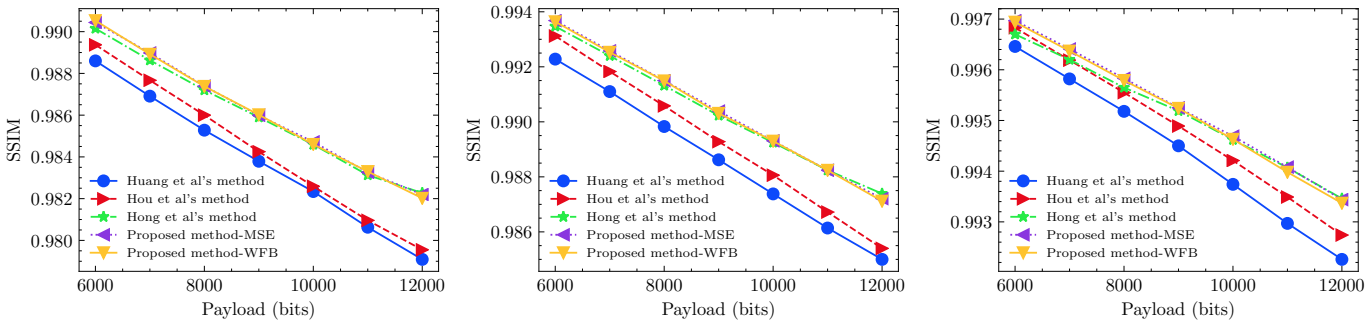
The results are shown in Fig. 9. Using $K = 2$ and $R = 0.6$ outperform others with respect to different quality factors: 70, 80 and 90. In the subsequent experiments, we use $K = 2, R = 0.6$ as the default parameters.

B. Visual Quality

For visual quality, different JPEG compression quality factors, i.e. QF=70, 80 and 90 are tested on the 100 images selected from BOSSbase 1.01 and the parameter setting is using $R = 0.6$ and $K = 2$. The message length ranges from 6000 to 12000 bits. To point out, the number of DCT coefficients meeting the requirement of some images is very small, so these images cannot finish the embedding process. To ensure fairness, we selected images that all algorithms could embed successfully for performance evaluation.

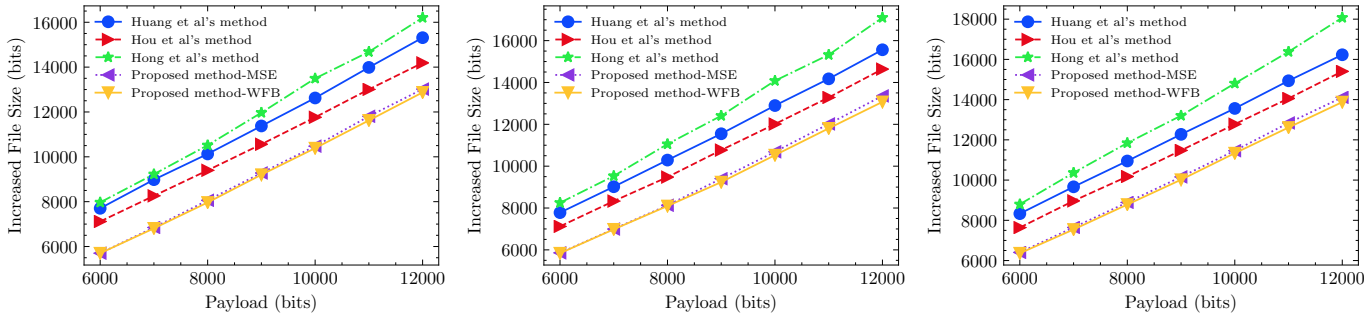
The PSNR and SSIM are calculated between the spatial domain images of the original JPEG file and the marked JPEG file. Various lengths of the message are exploited to show the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2020.3044186, IEEE Transactions on Circuits and Systems for Video Technology



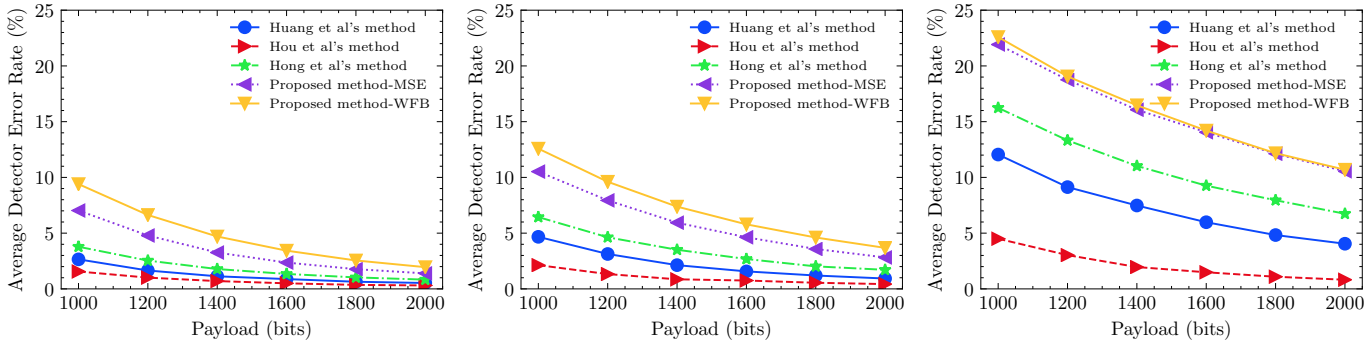
(a) QF = 70 (b) QF = 80 (c) QF = 90

Figure 11. The average **SSIM** values corresponding to different embedding payloads.



(a) QF = 70 (b) QF = 80 (c) QF = 90

Figure 12. The average **increased file** sizes corresponding to different embedding payloads.



(a) QF = 70 (b) QF = 80 (c) QF = 90

Figure 13. The average detection error rates of different hiding schemes against **DCTR**.

priority of our proposed schemes. The results are shown in Fig. 10 and Fig. 11, where the horizontal axes represent the embedding payloads and the vertical axes represent the average PSNR values and SSIM values.

It can be observed in Fig. 10, the average PSNR values obtained by the proposed methods (WFB and MSE) are larger than those obtained by other methods by a clear margin. As for SSIM in Fig. 11, the proposed methods are superior to other methods in most cases. These results imply that the proposed algorithm can achieve better visual quality in general.

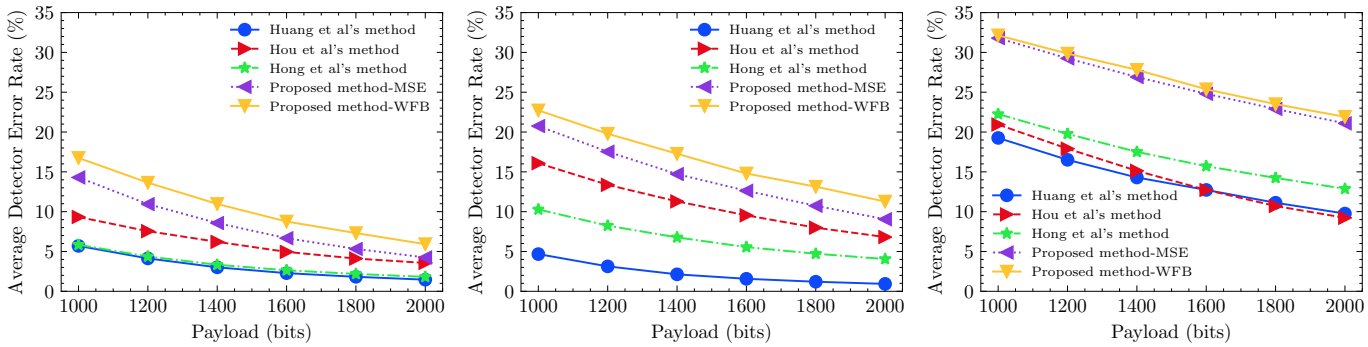
C. File Size Preservation

Fig. 12 shows the average increased file sizes between the original image and the marked image. The horizontal axes represent the embedding payloads and the vertical axes represent the increased file sizes (bits). The corresponding image quality factor is shown in the title of each subfigure as well. It is observed from Fig. 12 that the proposed method can preserve the file size better than other methods for different payloads and different quality factors.

D. Steganalytic Results

The extra profit of the proposed method is that the marked images own stronger undetectability than previous works, result-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2020.3044186, IEEE Transactions on Circuits and Systems for Video Technology



(a) QF = 70 (b) QF = 80 (c) QF = 90

Figure 14. The average detection error rates of different hiding schemes against **GFR**.

ing from the multi-distortion setting. Consequently, steganalytic experiments are carried out to verify the improvements. We select 10,000 images from the Alaska dataset, and transform them into JPEG images as the testing images using *imwrite* in MATLAB with quality factors 70, 80 and 90. Payloads range from 1,000 bits to 2,000 bits by step 200 bits. Analogously, three methods are compared. Since the modification is not continuous, the first-order statistical tests (e.g. Dual Statistic Test) [42]–[45] is hard to detect stego images. Instead, high-dimensional statistical features DCTR [33] and GFR [34] are selected as the steganalytic features, which own strong detection ability. The ensemble classifier with the default setting is utilized and output the detection error $P_{\mathrm{E}} = \min_{P_{\mathrm{FA}}} \frac{1}{2}(P_{\mathrm{FA}} + P_{\mathrm{MD}})$, where P_{FA} and P_{MD} are the false-alarm probability and the missed-detection probability respectively. The ultimate security is measured by average error rate $\overline{P}_{\mathrm{E}}$ over ten 50/50 database splits, and larger $\overline{P}_{\mathrm{E}}$ leads to stronger security.

The results are shown in Fig. 13 and Fig. 14. We can observe that the proposed schemes have a higher level of statistical undetectability with respect to different quality factors and message length. When the quality factor is large, the hiding method is more difficult to de detected. The distortion WFB performs better than MSE distortion, which shows that modifying coefficients causing less filter residual difference owns better security.

E. Time Complexity

We randomly select 1000 images to measure the running time of the mentioned methods at 6000 bits payload with different quality factors. The time measurement is performed with Matlab 2017a on a 3.20 GHz Intel Core i5 desktop computer with 8GB of memory running a 64-bit Ubuntu. As shown in Fig. 15, the computational time of Huang *et al.*'s method [15] is the shortest. Hou *et al.*'s method [9] is most time-consuming. Our proposed methods are competitive among these methods in terms of running speed.

F. Ablation Study

In this subsection, we explore the contribution of different parts in the proposed method, including selecting $\pm 1, \pm 2$ as the cover coefficients and the M-RHM scheme. Here, MSE

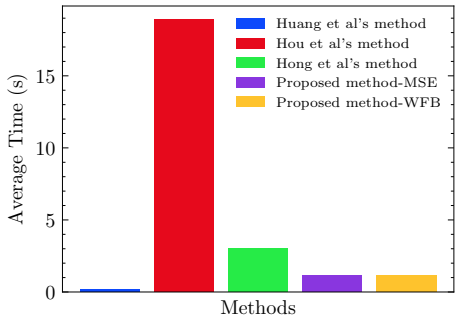


Figure 15. The running time of different methods with different quality factors.

is selected as the distortion and the results are shown in Fig. 16. $\pm 1, k = 1$ means selecting ± 1 as the cover coefficients and using consistent distortion for embedding. $\pm 1, \pm 2, k = 1$ means selecting $\pm 1, \pm 2$ as cover coefficients. $\pm 1, \pm 2, k = 2$ means using both strategies in the proposed method. The results show that selecting $\pm 1, \pm 2$ coefficients improve the visual quality, and M-RHM further enhances the performance.

V. Conclusions

In this paper, we present a novel JPEG reversible data hiding scheme using Recursive Histogram Modification under Multi-distortion metric (M-RHM), which consists of four modules: defining modification distortion, selecting suitable coefficients, embedding message, generating the final marked image. Benefited from the multi-distortion and selecting suitable coefficients, the visual quality and the ability of preserving file size outperform the existing methods. In addition, the proposed method owns superior undetectability.

The total distortion of modification is expressed in the additive form as the sum of embedding distortion over all elements. Actually, the modification will mutually interact, e.g. two coefficients modified simultaneously will cause smaller distortion than the sum of their single modifying distortion. In our future work, we will exploit the phenomenon, and propose non-additive distortion for JPEG RDH.

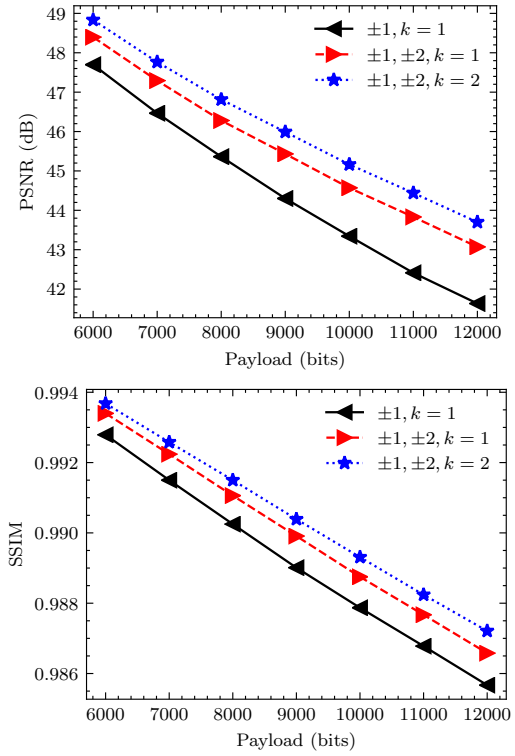


Figure 16. The visual quality of different schemes when QF=80.

Acknowledgment

To help readers apply the proposed method, we will post the MATLAB implementation of this paper at https://github.com/everange-ustc/JPEG-RDH

References

[1] J. M. Barton, "Method and apparatus for embedding authentication information within digital data," Jul. 8 1997, US Patent 5,646,997.

[2] J. Fridrich, M. Goljan, and R. Du, "Lossless data embedding for all image formats," in *Security and Watermarking of Multimedia Contents IV*, vol. 4675. International Society for Optics and Photonics, 2002, pp. 572–584.

[3] M. U. Celik, G. Sharma, A. M. Tekalp, and E. Saber, "Lossless generalized-LSB data embedding," *IEEE Transactions on Image Processing*, vol. 14, no. 2, pp. 253–266, 2005.

[4] A. M. Alattar, "Reversible watermark using the difference expansion of a generalized integer transform," *IEEE Transactions on Image Processing*, vol. 13, no. 8, pp. 1147–1156, 2004.

[5] D. M. Thodi and J. J. Rodríguez, "Expansion embedding techniques for reversible watermarking," *IEEE Transactions on Image Processing*, vol. 16, no. 3, pp. 721–730, 2007.

[6] Z. Ni, Y.-Q. Shi, N. Ansari, and W. Su, "Reversible data hiding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 3, pp. 354–362, 2006.

[7] S.-K. Lee, Y.-H. Suh, and Y.-S. Ho, "Reversible image authentication based on watermarking," in *International Conference on Multimedia and Expo*. IEEE, 2006, pp. 1321–1324.

[8] X. Li, B. Li, B. Yang, and T. Zeng, "General framework to histogram-shifting-based reversible data hiding," *IEEE Transactions on Image Processing*, vol. 22, no. 6, pp. 2181–2191, 2013.

[9] D. Hou, H. Wang, W. Zhang, and N. Yu, "Reversible data hiding in JPEG image based on DCT frequency and block selection," *Signal Processing*, vol. 148, pp. 41–47, 2018.

[10] G. Xuan, Y. Q. Shi, Z. Ni, P. Chai, X. Cui, and X. Tong, "Reversible data hiding for JPEG images based on histogram pairs," in *International Conference Image Analysis and Recognition*. Springer, 2007, pp. 715–727.

[11] H. Sakai, M. Kuribayashi, and M. Morii, "Adaptive reversible data hiding for JPEG images," in *International Symposium on Information Theory and Its Applications*. IEEE, 2008, pp. 1–6.

[12] Q. Li, Y. Wu, and F. Bao, "A reversible data hiding scheme for JPEG images," in *Pacific-Rim Conference on Multimedia*. Springer, 2010, pp. 653–664.

[13] T. Efimushkina, K. Egiazarian, and M. Gabbouj, "Rate-distortion based reversible watermarking for JPEG images with quality factors selection," in *European Workshop on Visual Information Processing*. IEEE, 2013, pp. 94–99.

[14] A. Nikolaidis, "Reversible data hiding in JPEG images utilising zero quantised coefficients," *IET Image Processing*, vol. 9, no. 7, pp. 560–568, 2015.

[15] F. Huang, X. Qu, H. J. Kim, and J. Huang, "Reversible data hiding in JPEG images," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 9, pp. 1610–1621, 2016.

[16] C.-C. Chang, C.-C. Lin, C.-S. Tseng, and W.-L. Tai, "Reversible hiding in DCT-based compressed images," *Information Sciences*, vol. 177, no. 13, pp. 2768–2786, 2007.

[17] C.-C. Lin and P.-F. Shiu, "DCT-based reversible data hiding scheme," in *International Conference on Ubiquitous Information Management and Communication*. ACM, 2009, pp. 327–335.

[18] L. S.-T. Chen, S.-J. Lin, and J.-C. Lin, "Reversible JPEG-based hiding method with high hiding-ratio," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 24, no. 03, pp. 433–456, 2010.

[19] K. Wang, Z.-M. Lu, and Y.-J. Hu, "A high capacity lossless data hiding scheme for JPEG images," *Journal of Systems and Software*, vol. 86, no. 7, pp. 1965–1975, 2013.

[20] B. G. Mobasseri, R. J. Berger, M. P. Marcinak, Y. J. NaikRaikar *et al.*, "Data embedding in JPEG bitstream by code mapping," *IEEE Transactions on Image Processing*, vol. 19, no. 4, pp. 958–966, 2010.

[21] Y. Wu and R. H. Deng, "Zero-error watermarking on JPEG images by shuffling huffman tree nodes," in *Visual Communications and Image Processing*. IEEE, 2011, pp. 1–4.

[22] Z. Qian and X. Zhang, "Lossless data hiding in JPEG bitstream," *Journal of Systems and Software*, vol. 85, no. 2, pp. 309–313, 2012.

[23] Y. Hu, K. Wang, and Z.-M. Lu, "An improved VLC-based lossless data hiding scheme for JPEG images," *Journal of Systems and Software*, vol. 86, no. 8, pp. 2166–2173, 2013.

[24] Z. Qian, H. Xu, X. Luo, and X. Zhang, "New framework of reversible data hiding in encrypted JPEG bitstreams," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 2, pp. 351–362, 2018.

[25] D. Hou, W. Zhang, Y. Yang, and N. Yu, "Reversible data hiding under inconsistent distortion metrics," *IEEE Transactions on Image Processing*, vol. 27, no. 10, pp. 5087–5099, 2018.

[26] Z. Hong, Z. Yin, and B. Luo, "Improved reversible data hiding in JPEG images based on interval correlation," in *International Conference on Brain Inspired Cognitive Systems*. Springer, 2018, pp. 687–696.

[27] W. Zhang, X. Hu, X. Li, and N. Yu, "Optimal transition probability of reversible data hiding for general distortion metrics and its applications," *IEEE Transactions on Image Processing*, vol. 24, no. 1, pp. 294–304, Jan 2015.

[28] ——, "Recursive histogram modification: establishing equivalency between reversible data hiding and lossless data compression," *IEEE Transactions on Image Processing*, vol. 22, no. 7, pp. 2775–2785, 2013.

[29] R. C. Gonzalez and R. E. Woods, "Digital image processing," *Nueva Jersey*, 2008.

[30] E. Hamilton, "JPEG file interchange format," 2004.

[31] C.-C. Chang and C.-T. Li, "Reversible data hiding in JPEG images based on adjustable padding," in *International Workshop on Biometrics and Forensics*. IEEE, 2017, pp. 1–6.

[32] A. Ichigaya, M. Kurozumi, N. Hara, Y. Nishida, and E. Nakasu, "A method of estimating coding PSNR using quantized DCT coefficients," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 2, pp. 251–259, 2006.

[33] V. Holub and J. Fridrich, "Low-complexity features for JPEG steganalysis using undecimated DCT," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 2, pp. 219–228, 2015.

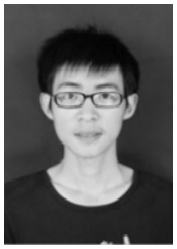
[34] X. Song, F. Liu, C. Yang, X. Luo, and Y. Zhang, "Steganalysis of adaptive JPEG steganography using 2D Gabor filters," in *Proceedings of the 3rd ACM Workshop on Information Hiding and Multimedia Security*. ACM, 2015, pp. 15–23.

[35] I. Daubechies and B. J. Bates, "Ten lectures on wavelets," 1993.

[36] B. Li, S. Tan, M. Wang, and J. Huang, "Investigation on cost assignment in spatial image steganography," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 8, pp. 1264–1277, 2014.

[37] L. Guo, J. Ni, and Y. Q. Shi, "Uniform embedding for efficient JPEG steganography," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 5, pp. 814–825, 2014.
[38] V. Holub and J. Fridrich, "Designing steganographic distortion using directional filters," in *International Workshop on Information Forensics and Security*. IEEE, 2012, pp. 234–239.
[39] R. Cogranne, Q. Giboulot, and P. Bas, "The alaska steganalysis challenge: A first step towards steganalysis," in *ACM Workshop on Information Hiding and Multimedia Security*, 2019, pp. 125–137.
[40] P. Bas, T. Filler, and T. Pevnỳ, "Break our steganographic system," in *International Workshop on Information Hiding*. Springer, 2011, pp. 59–70.
[41] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Asilomar Conference on Signals, Systems & Computers*, vol. 2. IEEE, 2003, pp. 1398–1402.
[42] A. Westfeld and A. Pfitzmann, "Attacks on steganographic systems," in *International Workshop on Information Hiding*, vol. 1768. Springer, 1999, pp. 61–76.
[43] N. Provos, "Defending against statistical steganalysis," in *Usenix security symposium*, vol. 10, 2001, pp. 323–336.
[44] J. Fridrich and M. Goljan, "Practical steganalysis of digital images: state of the art," in *Security and Watermarking of Multimedia*, vol. 4675. International Society for Optics and Photonics, 2002, pp. 1–13.
[45] G. Paul, I. Davidson, I. Mukherjee, and S. Ravi, "Keyless dynamic optimal multi-bit image steganography using energetic pixels," *Multimedia Tools and Applications*, vol. 76, no. 5, pp. 7445–7471, 2017.

Kejiang Chen received his B.S. degree in 2015 from Shanghai University (SHU) and a Ph.D. degree in 2020 from the University of Science and Technology of China (USTC). Currently, he is a postdoctoral researcher at the University of Science and Technology of China. His research interests include information hiding, image processing and deep learning.



Hang Zhou received his B.S. degree in 2015 from Shanghai University (SHU) and a Ph.D. degree in 2020 from the University of Science and Technology of China (USTC). His research interests include computer graphics, multimedia security and deep learning.

Dongdong Hou received his B.S. degree in 2014 from Hefei University of Technology (HUFT) and a Ph.D. degree in 2019 from the University of Science and Technology of China (USTC). His research interests include multimedia security and deep learning.

Weiming Zhang received his M.S. degree and Ph.D. degree in 2002 and 2005, respectively, from the Zhengzhou Information Science and Technology Institute, P.R. China. Currently, he is a professor with the School of Information Science and Technology, University of Science and Technology of China. His research interests include information hiding and multimedia security.

Nenghai Yu received his B.S. degree in 1987 from Nanjing University of Posts and Telecommunications, an M.E. degree in 1992 from Tsinghua University and a Ph.D. degree in 2004 from the University of Science and Technology of China, where he is currently a professor. His research interests include multimedia security, multimedia information retrieval, video processing and information hiding.