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.2020.3027391, IEEE Transactions on Circuits and Systems for Video Technology

Efficient reversible data hiding for JPEG images with multiple histograms modification

Mengyao Xiao, Xiaolong Li, Bin Ma, Xinpeng Zhang, and Yao Zhao

***Abstract*—Most current reversible data hiding (RDH) techniques are designed for uncompressed images. However, JPEG images are more commonly used in our daily lives. Up to now, several RDH methods for JPEG images have been proposed, yet few of them investigated the adaptive data embedding as the lack of accurate measurement for the embedding distortion. To realize adaptive embedding and optimize the embedding performance, in this paper, a novel RDH scheme for JPEG images based on multiple histogram modification (MHM) and rate-distortion optimization is proposed. Firstly, with selected coefficients, the RDH for JPEG images is generalized into a MHM embedding framework. Then, by estimating the embedding distortion, the rate-distortion model is formulated, so that the expansion bins can be adaptively determined for different histograms and images. Finally, to optimize the embedding performance in real time, a greedy algorithm with low computation complexity is proposed to derive the nearly optimal embedding efficiently. Experiments show that the proposed method can yield better embedding performance compared with state-of-the-art methods in terms of both visual quality and file size preservation.**

***Index Terms*—Reversible data hiding, JPEG image, MHM, rate-distortion optimization model, greedy algorithm.**

I. Introduction

As a special data hiding technique, reversible data hiding (RDH) has aroused much attentions in recent years. By RDH, not only the hidden data but also the original cover medium can be completely recovered by the authorized users [1]. Existing RDH methods are mainly designed for uncompressed images, and can be separated into four main categories: (1) compression based methods [2]–[4], (2) difference expansion based methods [5]–[9], (3) histogram shifting (HS) based methods [10]–[22], and (4) integer-to-integer transform based methods [23], [24]. However, up to now, for the most commonly used JPEG images, only a few RDH methods have been proposed. On the one hand, compared with the uncompressed format, JPEG converts the image from spatial domain into frequency domain and discards the high frequency information. The information of image is reduced, making it difficult to exploit the image redundancy for efficient RDH. On the other hand, prior knowledge about the uncompressed image fail to capture the effects of quantization and entropy encoding in JPEG, and therefore may not be useful for JPEG images RDH. Moreover, since the data embedding will generally increase the file size of the marked JPEG image, in addition to the conventional evaluation criterion, RDH for JPEG images also needs to consider the file size preservation. Keeping the increase of file size as little as possible after data embedding is an additional criterion to measure the embedding performance in this field.

As shown in Fig. 1, based on the compression process, existing RDH methods for JPEG images can be classified into three categories, i.e., the methods that modify the quantization table [25], [26], the methods that operate the bitstream [27]–[33], and the methods that modify DCT coefficients [22], [34]–[47].

For the first category, Fridrich *et al.* [25] firstly proposed to modify the quantization table for JPEG images RDH, in which some elements of the quantization table are divided by the integer 2 and the corresponding quantized DCT coefficients are simply multiplied by the same integer to make space for reversible embedding. Wang *et al.* [26] improved [25] by dividing some selected elements of quantization table with an integer larger than 2 so as to achieve a larger embedding capacity. Although this category of methods can well preserve the marked image quality, they also increase the file size significantly.

For the second category, Mobasseri *et al.* [27] used the JPEG stream to embed data by mapping the used variable length code (VLC) to an unused VLC, thus the file size of marked image is well preserved. This method was further improved by Qian and Zhang [28] and Hu *et al.* [29], where the unused codes have been fully exploited to design better mapping strategy, and higher embedding capacity is achieved. Moreover, in [30], Qian *et al.* developed a JPEG images RDH method in encrypted domain, in which the JPEG bitstream is encrypted into an organized structure, and the secret data is embedded into the encrypted bitstream by modifying the JPEG stream. Generally, this category of methods can well preserve the file size while ensure the image fidelity. However, the main drawback of these methods is that the embedding capacity is rather low.

The third category, which is the most popular approach, mainly concentrates in the modification of the quantized DCT coefficients. In [34], Fridrich *et al.* first proposed to lossless

Corresponding author: Yao Zhao

M. Xiao, X. Li and Y. Zhao are with the Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China, and also with Beijing Key Laboratory of Advanced Information Science and Network Technology, Beijing 100044, China (e-mail: xiaomengyao@bjtu.edu.cn; lixl@bjtu.edu.cn; yzhao@bjtu.edu.cn).

B. Ma is with the School of Cyber Security, Qilu University of Technology (Shandong Academy of Sciences), Jinan 250353, China, and also with the Shandong Provincial Key Laboratory of Computer Networks, Jinan 250353, China (e-mail: sddxmb@126.com).

X. Zhang is with the School of Computer Science, Fudan University, Shanghai 200433, China (e-mail: zhangxinpeng@fudan.edu.cn).

This work was supported by the National Key Research and Development of China (No. 2018AAA0102100), the National Natural Science Foundation of China (Nos. 61972031, U1936212 and U1736213), and the Fundamental Research Funds for the Central Universities (No. 2018JBZ001 and 2019YJS028).

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.2020.3027391, IEEE Transactions on Circuits and Systems for Video Technology

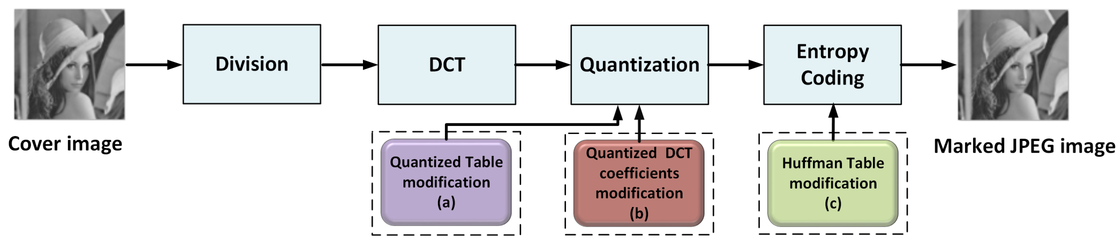

Fig. 1. Three different types of RDH methods for JPEG images (a) the methods that modify the quantization table (b) the methods that modify quantized DCT coefficients (c) the methods that modify the Huffman table.

compress the LSB plane of the quantized DCT coefficients so as to create space for reversible embedding. Chang *et al.* [35] proposed to modify the successive zero valued coefficients in the middle frequency and cut off the low frequency components as humans are more sensitive to the changes in these frequencies. Xuan *et al.* [36] proposed to apply the HS technique for JPEG images RDH, in which multiple pairs of expansion bins in the histogram of quantized DCT coefficients are utilized to embed data. Sakai *et al.* [37] improved Xuan *et al.*'s scheme by selecting the smooth 8×8 DCT blocks for data embedding, where the smoothness is evaluated by the fluctuation of the direct current (DC) coefficients in its neighboring blocks. Efimushkina *et al.* [38] proposed to select coefficients with small magnitudes for data embedding so as to introduce less distortion. In [40], Huang *et al.* proposed to use the alternating current (AC) coefficients valued 1 and -1 for expansion embedding while remain the zero valued AC coefficients unchanged to preserve the file size. Besides, a DCT block selection strategy is utilized to embed the smooth blocks preferentially. As an extension of Huang *et al.*'s work, Wedaj *et al.* [41] proposed to calculate the embedding efficiency for different DCT bands and then select the bands with higher efficiency to embed data. Similarly, Hou *et al.* [43] proposed to calculate the unit distortion for different DCT bands and then select bands with less unit distortion for data embedding. In [45], He *et al.* proposed to establish the negative influence models of image visual distortion and file size change for band selection. Moreover, the secret data is embedded into the AC coefficients in an ascending order of zero run length. In a recent work [46], Yin *et al.* proposed a multi-objective optimization strategy. They considered both the image quality and the file size increase. Compared with the previous works, the third category of methods provide larger embedding capacity, and can guarantee image quality as well as the file size preservation.

This paper addresses JPEG images RDH based on quantized DCT coefficients modification. Notice that, for most previous methods of this type, the expansion coefficients are determined by some empirical criteria without considering the image content. The methods with additional band and block selection generally yield better embedding performance. However, due to the lack of accurate estimation for the embedding distortion, the performance is still far from optimal. Then, to realize adaptive embedding and optimize the embedding performance, in this paper, the RDH for JPEG images is generalized into a multiple histograms modification (MHM) framework, and the rate-distortion model is formulated to adaptive determine the optimal expansion bins for different histograms. Firstly, the AC coefficients in each DCT band are counted to generate 63 histograms, and the MHM-based reversible embedding framework is established by modifying each histogram individually. Then, by estimating the embedding distortion, the corresponding rate-distortion model is formulated based on MHM. Finally, to optimize the embedding performance in real time, a greedy algorithm with low computation complexity is proposed to derive the nearly optimal embedding efficiently. Experimental results show that with a good file size preservation, the proposed method can achieve better visual quality compared to some state-of-the-art methods [40], [41], [43].

The remainder of the paper is organized as follows. Section II introduces two state-of-the-art methods on this issue. Section III presents the proposed method. The experimental results are discussed in Section IV, and the conclusions are given in Section V.

II. Preliminary

In this section, two state-of-the-art JPEG images RDH methods [40] and [41] are reviewed. Firstly, as a preparation, the JPEG compression process is introduced as below.

As shown in Fig. 2, in JPEG compression, an uncompressed image is first divided into non-overlapping blocks of 8×8 pixels. Denote the blocks as $X_1, ..., X_N$, where N is the total number of divided blocks. For a given block X_k, suppose that

$$X_k = \begin{bmatrix} x_{k,0} & x_{k,1} & \cdots & x_{k,28} \\ x_{k,2} & \ddots & \ddots & x_{k,42} \\ \vdots & \ddots & \ddots & \vdots \\ x_{k,35} & x_{k,36} & \cdots & x_{k,63} \end{bmatrix} \quad (1)$$

in which the 64 pixels are indexed in a zig-zag order for simplicity. In this way, each block X_k is a vector with length 64. Then, DCT is applied to each block X_k, and correspondingly, the transformed block denoted $Y_k = (y_{k,i})_{i=0}^{63}$ is derived. Next, by a given quantization table $Q = (q_i)_{i=0}^{63}$, Y_k is quantized to yield the quantized block $Z_k = (z_{k,i})_{i=0}^{63}$, i.e., for each $0 \leq i \leq 63$, the quantized coefficient $z_{k,i}$ is defined by

$$z_{k,i} = \left[\frac{y_{k,i}}{q_i} \right] \quad (2)$$

where $[\cdot]$ is the round function. Here, $z_{k,0}$ is the DC coefficient, and $z_{k,i}$ for $i \in \{1, ..., 63\}$ are the AC coefficients. Finally, these quantized coefficients are encoded to obtain a compressed bitstream, and the JPEG file is generated.

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.2020.3027391, IEEE Transactions on Circuits and Systems for Video Technology

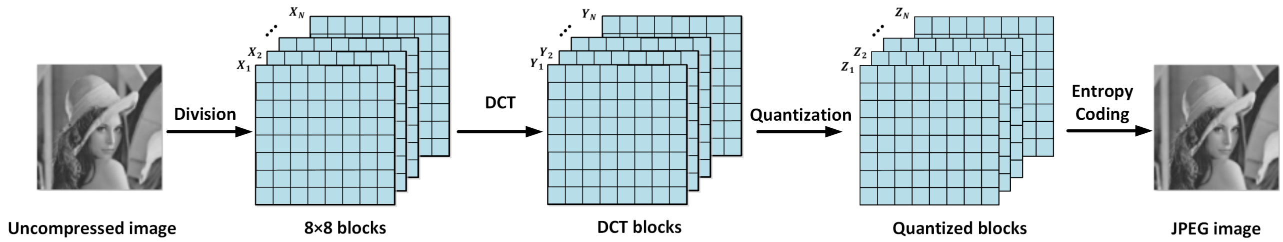

Fig. 2. JPEG compression process.

For most JPEG images RDH methods including [40] and [41], the data embedding is conducted by modifying the AC coefficients while the DC coefficients are unchanged.

A. Huang et al.'s Method [40]

In Huang *et al.*'s method [40], the AC coefficients valued 1 and -1 in some selected smooth blocks are utilized for expansion embedding. First, for each block, the number of zero valued AC coefficients is counted as its smoothness measurement, i.e., the smoothness for the k-th block is defined as

$$T_k = \#\{i : z_{k,i} = 0,\ 1 \leq i \leq 63\} \tag{3}$$

where $\#$ means the cardinality of a set. Then, based on the smoothness, the quantized blocks sequence $(Z_1, ..., Z_N)$ is sorted in descending order to generate a new sequence $(Z_{\alpha(1)}, ..., Z_{\alpha(N)})$, where $\alpha : \{1, ..., N\} \to \{1, ..., N\}$ is the unique one-to-one mapping such that $T_{\alpha(1)} \geq ... \geq T_{\alpha(N)}$, and $\alpha(i) < \alpha(j)$ if $T_{\alpha(i)} = T_{\alpha(j)}$ with $i < j$. Finally, for a given p, the AC coefficients in the first p blocks are modified to embed data, i.e., for each $k \in \{\alpha(1), ..., \alpha(p)\}$ and $i \in \{1, ..., 63\}$, $z_{k,i}$ is modified to $z^*_{k,i}$ as

$$z^*_{k,i} = \begin{cases} z_{k,i}, & \text{if } z_{k,i} = 0 \\ z_{k,i} + m, & \text{if } z_{k,i} = 1 \\ z_{k,i} - m, & \text{if } z_{k,i} = -1 \\ z_{k,i} + 1, & \text{if } z_{k,i} > 1 \\ z_{k,i} - 1, & \text{if } z_{k,i} < -1 \end{cases} \tag{4}$$

where $m \in \{0, 1\}$ is a to-be-embedded bit. In this way, for the selected blocks, the AC coefficients valued 1 and -1 are expanded to carry data while the other non-zero ones are shifted to create the vacancy. As the zero valued AC coefficients remain unchanged after data embedding, the same T_k can be derived by both the encoder and decoder, and thus the reversibility is guaranteed. The modified quantized block is denoted as Z^*_k for each $1 \leq k \leq N$. Notice that, in the above process, only the first p smooth blocks are used for data embedding, where the rough blocks with indices $\alpha(p + 1), ..., \alpha(N)$ are unchanged. And, the selected block number p is determined as the smallest one that can provide sufficient embedding capacity.

For data extraction, the smoothness T_k of each block is calculated firstly. Then, for each smooth block Z^*_k with $k \in \{\alpha(1), ..., \alpha(p)\}$, the modified AC coefficients are restored as

$$z_{k,i} = \begin{cases} 0, & \text{if } z^*_{k,i} = 0 \\ 1, & \text{if } 1 \leq z^*_{k,i} \leq 2 \\ -1, & \text{if } -2 \leq z^*_{k,i} \leq -1 \\ z^*_{k,i} - 1, & \text{if } z^*_{k,i} > 2 \\ z^*_{k,i} + 1, & \text{if } z^*_{k,i} < -2 \end{cases} \tag{5}$$

and the embedded data bits are extracted as

$$m = \begin{cases} 0, & \text{if } |z^*_{k,i}| = 1 \\ 1, & \text{if } |z^*_{k,i}| = 2 \end{cases}. \tag{6}$$

The DC coefficients and the rough blocks are unmodified during data embedding, they are recovered as themselves.

Experimental results reported in [40] have shown that this method outperforms the previous works such as [36] and [37] in terms of both visual quality and file size preservation.

Finally, we remark that, in the viewpoint of histogram modification, this method can be introduced equivalently based on HS. Actually, a histogram denoted h is defined as follows to count the occurrences of all AC coefficients in the selected smooth blocks, i.e., for each $s \in \mathbb{Z}$,

$$h(s) \triangleq \#\{(k, i) : z_{\alpha(k),i} = s,\ 1 \leq k \leq p,\ 1 \leq i \leq 63\}. \tag{7}$$

Then, this histogram is modified in which the bins 1 and -1 are expanded to carry data, the bins larger than 1 or smaller than -1 are shifted to guarantee the reversibility, while the bin 0 is unchanged.

B. Wedaj et al.'s Method [41]

The embedding rule of Wedaj *et al.*'s method [41] is just the same as [40], i.e., using (4). However, for the selected blocks, instead of utilizing all the AC coefficients, the data embedding in [41] is restricted on some specific AC bands.

For this method, first, the quantized blocks sequence $(Z_1, ..., Z_N)$ is sorted into $(Z_{\alpha(1)}, ..., Z_{\alpha(N)})$ in the same way as [40]. Then, for each AC band $i \in \{1, ..., 63\}$, the embedding efficiency denoted R_i is calculated as

$$R_i = \frac{\sum_{k=1}^{N} E_{k,i}}{q_i^2 \sum_{k=1}^{N} (S_{k,i} + \frac{E_{k,i}}{2})} \tag{8}$$

where $E_{k,i} \in \{0, 1\}$ indicates whether the AC coefficient $z_{k,i}$ is embeddable or not, i.e.,

$$E_{k,i} = \begin{cases} 1, & \text{if } |z_{k,i}| = 1 \\ 0, & \text{if } |z_{k,i}| \neq 1 \end{cases} \tag{9}$$

and $S_{k,i} \in \{0,1\}$ denotes whether $z_{k,i}$ is shiftable or not, i.e.,

$$S_{k,i} = \begin{cases} 1, & \text{if } |z_{k,i}| > 1 \\ 0, & \text{if } |z_{k,i}| \leq 1 \end{cases}. \tag{10}$$

By this definition, the numerator of R_i represents the embedding capacity for the i-th AC band while the denominator represents the corresponding estimated embedding distortion, and larger R_i means that less distortion will be introduced. Based on the values of R_i, all the AC bands are sorted in descending order to derive a new band sequence $(\beta(1), ..., \beta(63))$, where $\beta : \{1, ..., 63\} \to \{1, ..., 63\}$ is the unique one-to-one mapping such that $R_{\beta(1)} \geq ... \geq R_{\beta(63)}$, and $\beta(i) < \beta(j)$ if $R_{\beta(i)} = R_{\beta(j)}$ with $i < j$. Next, for given p and l, the AC coefficients of blocks $\{Z_{\alpha(1)}, ..., Z_{\alpha(p)}\}$ in the bands $\{\beta(1), ..., \beta(l)\}$ are counted to generate a histogram $h_{p,l}$ defined as

$$h_{p,l}(s) = \#\{(k,i) : z_{\alpha(k),\beta(i)} = s,\ 1 \leq k \leq p,\ 1 \leq i \leq l\}. \tag{11}$$

After that, for each $l \in \{1, ..., 63\}$, the selected block number p is determined as the smallest value such that the histogram $h_{p,l}$ can provide sufficient embedding capacity. Then, the band number is determined as the one that lead to the lowest distortion, and the data embedding is finally conducted by modifying the histogram $h_{p,l}$ with selected l and p. With the band selection strategy, less distortion is introduced in [41] compared with [40].

In this method, although the band selection strategy can provide embedding performance improvement, the performance is still far from optimal since there is no accurate estimate for the embedding distortion. Moreover, the expansion bins, i,e,. 1 and -1, are just empirically determined without considering the image content. Then, to realize adaptive embedding and optimize the embedding performance, in this paper, the RDH for JPEG images is generalized into a MHM framework, and the rate-distortion model is formulated to adaptive determine the optimal expansion bins for different histograms. The details are given in the next section.

III. Proposed Method

The previous methods have not studied the optimal embedding for JPEG images RDH as the lack of embedding distortion estimation. To optimize the embedding performance, a MHM-based RDH scheme for JPEG images is proposed in this section. Firstly, a MHM-based reversible embedding framework is established, and the corresponding rate-distortion model is formulated. Then, to optimize the embedding performance in real time, a greedy algorithm with low computation complexity is proposed in Section III-B. Finally, the implementation details of the proposed scheme are presented in Section III-C.

A. MHM Framework for JPEG Images RDH

Using the notations introduced in Section II, for selected AC coefficients $\{z_{k,i} : k \in S_i\}$, where S_i is a subset of $\{1, ..., N\}$, a histogram h_i is defined for each $i \in \{1, ..., 63\}$, i.e.,

$$h_i(s) = \#\{k : z_{k,i} = s \mid k \in S_i\}. \tag{12}$$

In this way, h_i counts the occurrences of AC coefficients in the i-th band. By this definition, 63 histograms are generated, and the data embedding is then conducted by modifying each histogram based on MHM. That is to say, for each h_i, one pair of bins $a_i < b_i$ is selected for expansion, in which the bins between a_i and b_i remain unchanged, and the bins smaller than a_i or larger than b_i are shifted for reversibility. Specifically, for each $k \in S_i$, the AC coefficient $z_{k,i}$ is modified to $z^*_{k,i}$ as

$$z^*_{k,i} = \begin{cases} z_{k,i}, & \text{if } a_i < z_{k,i} < b_i \\ z_{k,i} + m, & \text{if } z_{k,i} = b_i \\ z_{k,i} - m, & \text{if } z_{k,i} = a_i \\ z_{k,i} + 1, & \text{if } z_{k,i} > b_i \\ z_{k,i} - 1, & \text{if } z_{k,i} < a_i \end{cases} \tag{13}$$

where $m \in \{0,1\}$ is a to-be-embedded bit. Here, the other AC coefficients $z_{k,i}$ with $k \notin S_i$ are unchanged. Accordingly, at the extraction side, for each $k \in S_i$, the marked AC coefficient $z^*_{k,i}$ can be restored as

$$z_{k,i} = \begin{cases} z^*_{k,i}, & \text{if } a_i \leq z^*_{k,i} \leq b_i \\ z^*_{k,i} - 1, & \text{if } z^*_{k,i} > b_i \\ z^*_{k,i} + 1, & \text{if } z^*_{k,i} < a_i \end{cases} \tag{14}$$

and the embedded data can be extracted as $m = 0$ if $z^*_{k,i} \in \{a_i, b_i\}$ or $m = 1$ if $z^*_{k,i} \in \{a_i - 1, b_i + 1\}$.

Notice that, the previous methods [40] and [41] introduced in Section II are actually special cases of the above MHM-based embedding, in which the selected DCT blocks and expansion bins (a_i, b_i) are as follows:

- for [40]: $S_i = \{\alpha(1), ..., \alpha(p)\}$ and $(a_i, b_i) = (-1, 1)$ for each $i \in \{1, ..., 63\}$.
- for [41]: $S_i = \{\alpha(1), ..., \alpha(p)\}$ for each $i \in \{1, ..., 63\}$, $(a_i, b_i) = (-1, 1)$ if $i \in \{\beta(1), ..., \beta(p)\}$, and $(a_i, b_i) = (-\infty, +\infty)$ if $i \notin \{\beta(1), ..., \beta(p)\}$. Specifically, $(a_i, b_i) = (-\infty, +\infty)$ means that the histogram h_i is unchanged in data embedding.

For the above MHM-based embedding, we now discuss the rate-distortion model, upon which the embedding capacity and distortion can be formulated. Clearly, the embedding capacity noted EC can be calculated as

$$EC = \sum_{i=1}^{63} h_i(a_i) + h_i(b_i). \tag{15}$$

On the other hand, considering that the embedding distortion noted ED measures the degradation in spatial domain, we first introduce some notations as follows. As shown in Fig. 3, after data embedding, Z^*_k is multiplied by Q, point by point, to yield the dequantized coefficients $Y^*_k = (y^*_{k,i})_{0 \leq i \leq 63}$. And then the inverse DCT (IDCT) is performed on Y^*_k to obtain $X^*_k = (x^*_{k,i})_{0 \leq i \leq 63}$. Finally, X^*_k is further rounded and truncated in the range of $[0, 255]$ to get $\widetilde{X}_k = (\widetilde{x}_{k,i})_{0 \leq i \leq 63}$. Therefore, ED can be formulated as

$$ED = \sum_{k=1}^{N} \sum_{i=0}^{63} (\widetilde{x}_{k,i} - x_{k,i})^2. \tag{16}$$

Notice that, between $x^*_{k,i}$ and $\widetilde{x}_{k,i}$, there is a rounding and truncation operation. However, since the rounding error is very

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.2020.3027391, IEEE Transactions on Circuits and Systems for Video Technology

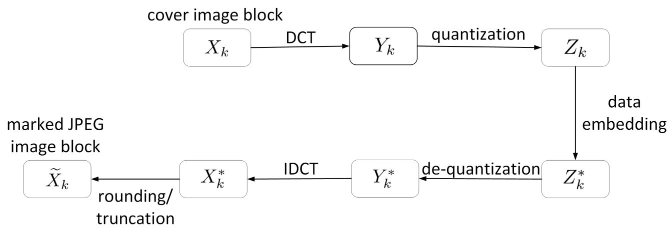

Fig. 3. Schematic diagram of DCT coefficients modification.

small which is an independent identically distributed random variable with uniform distribution in the range of $[-0.5, +0.5)$, and the truncation error occurs less than 1% in most cases [48], it actually leads to a negligible error. So, we can derive that

$$ED \approx \sum_{k=1}^{N}\sum_{i=0}^{63}(x_{k,i}^* - x_{k,i})^2. \tag{17}$$

In addition, as $Y_k = \text{DCT}(X_k)$ and $Y_k^* = \text{DCT}(X_k^*)$, then, $(Y_k^* - Y_k) = \text{DCT}(X_k^* - X_k)$. According to Parseval's theorem, we know that

$$\sum_{i=0}^{63}(x_{k,i}^* - x_{k,i})^2 = \sum_{i=0}^{63}(y_{k,i}^* - y_{k,i})^2. \tag{18}$$

Moreover, as $y_{k,i}^* = q_i z_{k,i}^*$ and $y_{k,i} \approx q_i z_{k,i}$, we have

$$\sum_{i=0}^{63}(y_{k,i}^* - y_{k,i})^2 \approx \sum_{i=0}^{63} q_i^2 (z_{k,i}^* - z_{k,i})^2. \tag{19}$$

Combined (17)-(19) and the fact that the DC coefficients are unchanged in data embedding, we can obtain that

$$ED \approx \sum_{k=1}^{N}\sum_{i=1}^{63} q_i^2 (z_{k,i}^* - z_{k,i})^2. \tag{20}$$

That is to say,

$$ED \approx \sum_{i=1}^{63} q_i^2 \sum_{k=1}^{N}(z_{k,i}^* - z_{k,i})^2 = \sum_{i=1}^{63} ED_i \tag{21}$$

where, for each $i \in \{1, ..., 63\}$

$$ED_i = q_i^2 \left(\frac{1}{2}(h_i(a_i) + h_i(b_i)) + \sum_{s<a_i} h_i(s) + \sum_{s>b_i} h_i(s) \right). \tag{22}$$

Therefore, based on (15) and (21), the rate-distortion optimization problem for payload-limited reversible embedding can be formulated as

$$\begin{cases} \text{minimize} & \frac{\sum_{i=1}^{63} ED_i}{EC} \\ \text{subject to} & EC \geq P \end{cases} \tag{23}$$

where P is the given payload and a parameter set $\{(a_i, b_i) : 1 \leq i \leq 63\}$ is required to be determined.

Notice that, here we assume that the cover image is in uncompressed format. When the cover image is in compressed form with JPEG, the analysis is similar and the details are omitted here.

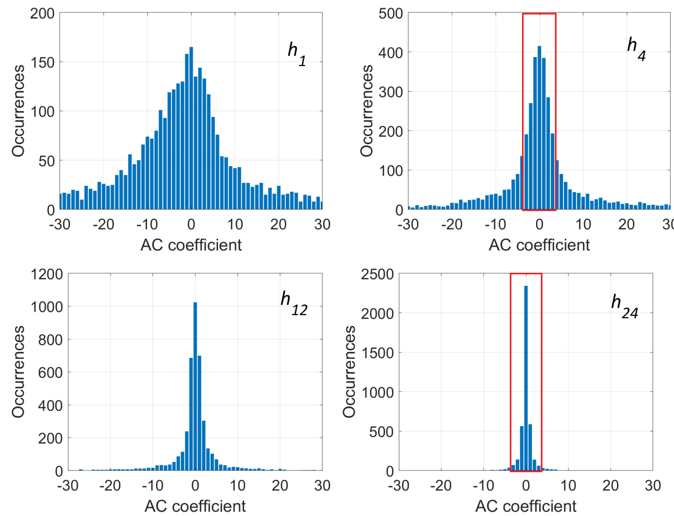

Fig. 4. The histograms h_1, h_4, h_{12} and h_{24} for the Lena image with QF = 90.

B. Optimal Parameters Determination of MHM

Based on (23), this subsection focus on the optimal parameters determination for MHM-based embedding.

Since the construction of the multiple histograms is completely different from the previous MHM [17], the parameters setting and determination here will be absolutely different. We first impose some constrains for the to-be-determined parameters $\{(a_i, b_i) : 1 \leq i \leq 63\}$. Notice that, for JPEG images, high frequency AC coefficients are usually quantized by a large step, and the modification on these coefficients generally results in more distortion compared with other frequencies. Then, the histograms for high frequency AC coefficients will not be used in the proposed data embedding. Specifically, considering $\Omega = \{36, ..., 63\}$ in the lower triangle of DCT block as high frequency AC bands, for each $i \in \Omega$, we take

$$(a_i, b_i) = (-\infty, +\infty). \tag{24}$$

For histograms of low and middle frequency, take h_1, h_4, h_{12}, h_{24} of Lena image with quality factor (QF) of 90 as an example (see Fig. 4). As shown, the histograms obey a Laplacian-like distribution centred at 0 and decay rapidly to 0 from both sides. Then, considering the symmetry of these histograms, for each band $i \in \{1, ..., 35\}$, we set

$$a_i = -b_i \quad \text{and} \quad b_i \in \{1, ..., M-1, +\infty\} \tag{25}$$

where M is a parameter controlling the choices of expansion bins. Here, the bin 0 will not be taken for expansion embedding, since the modification of zero valued AC coefficients usually leads to a significantly file size increase as pointed in [40].

Therefore, the actual variables for the optimization problem (23) is $(b_1, ..., b_{35})$ containing 35 parameters. Note that, as M increases, the number of candidates for $(b_1, ..., b_{35})$ increases exponentially, and the exhaustive search utilized in the previous MHM [17] is no longer applicable to find the optimal solution. For example, when $M = 6$, the computational complexity by using exhaustive search is about

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.2020.3027391, IEEE Transactions on Circuits and Systems for Video Technology

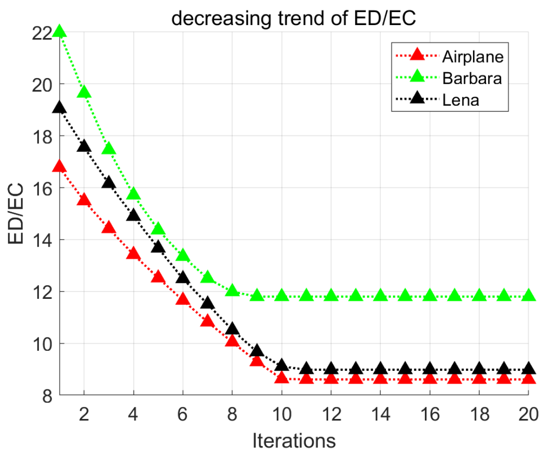

Fig. 5. The decreasing trend of $(\sum ED_i)/EC$ with respect to the number of iterations for Airplane, Barbara, and Lena images with capacities equal to 10000 bits.

$O(2^{91})$. Then, to solve the optimization problem, a greedy algorithm with low computation complexity is proposed here to derive the nearly optimal parameters efficiently. The main idea of the proposed greedy algorithm is to divide the optimizing process into multiple iterations. For each iteration, instead of considering all 35 parameters, only two parameters are optimized to minimize $(\sum_{i=1}^{35} ED_i)/EC$ based on (23) while other parameters keep unchanged. For details, first of all, for each h_i with $1 \leq i \leq 35$, initialize b_i as 1 to maximize the embedding capacity. Then, for each (i,j) with $1 \leq i < j \leq 35$, taking every choice of (b_i, b_j) while keeping other parameters unchanged. In this way, only $C_{35}^2 M^2$ cases need to be tested at most. Among all these cases, take the best one noted $(b'_1, ..., b'_{35})$ that minimize $(\sum_{i=1}^{35} ED_i)/EC$ while meeting the capacity requirement. After that, update the parameters $(b_1, ..., b_{35})$ by $(b'_1, ..., b'_{35})$, and one round iteration is completed. Repeat the above iterative process until the objective function $(\sum_{i=1}^{35} ED_i)/EC$ no longer reduces, and the nearly optimal parameters are finally obtained. For instance, Fig. 5 shows the decreasing trend of $(\sum ED_i)/EC$ with respect to the number of iterations for three commonly used images. Practically, this process will be stopped with at most 20 iterations according to our experiments.

The computation complexity of the proposed greedy algorithm is $O(C_{35}^2 M^2)$, which is much lower compared with $O(M^{35})$ by using exhaustive search. For example, when $M = 6$, the computational complexity is reduced from $O(2^{91})$ to $O(2^{15})$ by the proposed method. The greedy algorithm effectively solves the optimization problem, making it possible to find the nearly optimal parameters in a short time.

C. Implementation Details

We first introduce the coefficient selection strategy. By sorting the coefficients in a decreasing order of smoothness, the coefficients with small magnitudes are selected first to generate more concentrated histograms. On the one hand, as noted in [40], DCT blocks with more zero valued coefficients usually belong to the smoother image region, and the coefficients in these blocks tend to have small magnitudes. On the other hand, even with the same blocks, the sharpness of different histograms varies greatly, and the histograms with more zero coefficients usually have more proportion of small magnitude coefficients (see Fig. 4). For example, the coefficients with magnitudes less than 3 account for almost all the coefficients in h_{24}, while in h_4, these coefficients only account for 42%. Therefore, for the coefficient smoothness measurement, not only the number of zero valued coefficients for each block, but also the number of zero valued coefficients in each band for all blocks should be considered. Specifically, for each k, the smoothness noted TB_k of the k-th block is calculated as

$$TB_k = \#\{i : z_{k,i} = 0,\ 1 \leq i \leq 63\}. \tag{26}$$

Moreover, for each i, the number of zero valued coefficients in the i-th band denoted as F_i is calculated as

$$F_i = \#\{k : z_{k,i} = 0,\ 1 \leq k \leq N\}. \tag{27}$$

As the value of F_i is too large to combine with TB_k, instead of directly utilizing F_i, we consider taking a more appropriate value based on F_i, i.e., the rank of each F_i. That is to say, sort $(F_1, ..., F_{63})$ in ascending order to obtain the band ranking sequence $(\gamma(1), ..., \gamma(63))$, where $\gamma : \{1, ..., 63\} \rightarrow \{1, ..., 63\}$ is the unique one-to-one mapping such that $F_{\gamma(1)} \leq ... \leq F_{\gamma(63)}$ and $\gamma(i) < \gamma(j)$ if $F_{\gamma(i)} = F_{\gamma(j)}$ with $i < j$. The smoothness of the i-th band is then represented as $TF_i = \gamma^{-1}(i)$. Based on TB_k and TF_i, the smoothness measurement $T_{k,i}$ for the AC coefficient $z_{k,i}$ is finally set as

$$T_{k,i} = \lfloor aTF_i + (1-a)TB_k \rfloor \tag{28}$$

where $\lfloor \cdot \rfloor$ is the floor function, $0 \leq a \leq 1$ is a tunable parameter controlling the proportion of the two smoothness measurements. A large $T_{k,i}$ indicates that the coefficient is located in a smoother image region and should be used preferentially for data embedding.

The whole data embedding procedure of the proposed method is described as shown in Fig. 6. After determining all $T_{k,i}$ for $1 \leq k \leq N$ and $1 \leq i \leq 35$, a threshold $T \in \{0, ..., 63\}$ is set to select the smooth coefficients whose smoothness measurement is no less than T. Next, for each T, with the selected coefficients, solve the optimization problem (23) based on the proposed greedy algorithm, and compute the corresponding objective function $(\sum_{i=1}^{35} ED_i)/EC$. Finally, by testing every $T \in \{0, ..., 63\}$, the optimal T is determined as the one that the corresponding objective function $(\sum_{i=1}^{35} ED_i)/EC$ is minimized while providing sufficient payload. In this way, with the determined T, the coefficients in different bands are selected and the corresponding parameters $(b_1, ..., b_{35})$ are derived as well. With the selected coefficients and the derived parameters $(b_1, ..., b_{35})$, the data embedding process is then conducted according to (13).

To ensure the lossless recovery, the parameters $(b_1, ..., b_{35})$ with $35 \log_2 M$ bits and the smoothness threshold T with 6 bits should be taken as the side information, as well as the payload length (18 bits for a 512×512 image). The side information is totally $(35 \log_2 M + 24)$ bits and will be embedded into the 36th band (in zig-zag order) of the previous DCT blocks

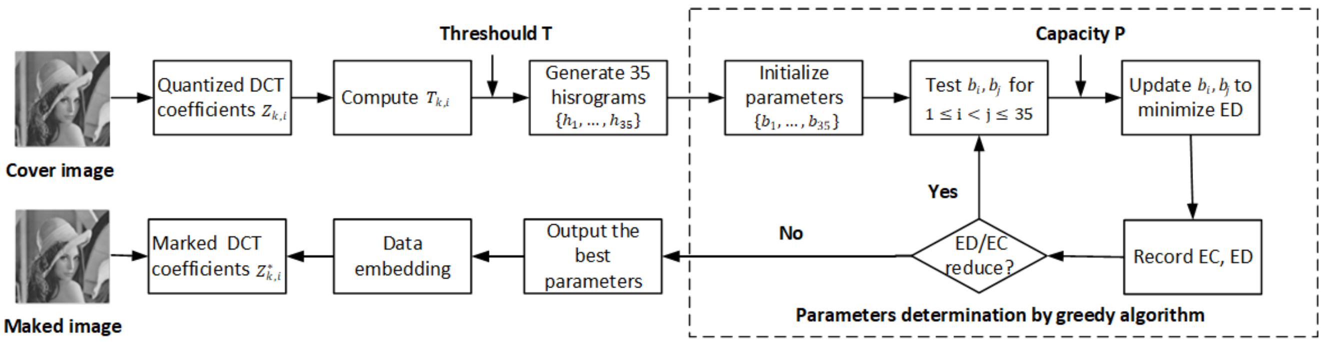

Fig. 6. Flowchart for the proposed method.

in order. Here, the data embedding for side information is processed just according to [40], i.e., only coefficients valued 1 and -1 are utilized for expansion.

For data extraction, firstly, with the obtained quantized DCT blocks from the decoder, the side information is extracted from the 36th band of the previous blocks in order. Then, $T_{k,i}$ for coefficients $z^*_{k,i}$ with $1 \le k \le N$ and $1 \le i \le 35$ are calculated in the same way as described above based on (28). Next, with the obtained T, $(b_1, \ldots, b_{35})$, and payload length, the marked AC coefficients $z^*_{k,i}$ are restored according to (14), and the embedded data can be extracted as $m = 0$ if $z^*_{k,i} \in \{a_i, b_i\}$ or $m = 1$ if $z^*_{k,i} \in \{a_i - 1, b_i + 1\}$. Finally, the recovered quantized DCT coefficients are entropy encoded to obtain the original image.

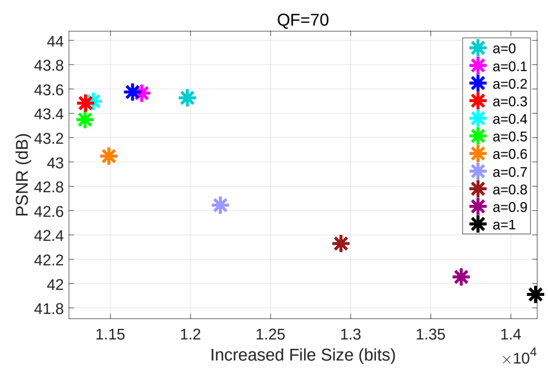

Fig. 7. The average embedding performance comparison for different a with eight images of QF $= 70$ by embedding 10,000 bits.

IV. Experimental Results and Discussions

In this section, the proposed method is evaluated by comparing it with three state-of-the-art methods, including Huang *et al.*'s method [40], Wedaj *et al.*'s method [41] and Hou *et al.*'s method [43]. The test images used in the experiments (including Airplane, Baboon, Barbara, Boat, Elaine, Lake, Lena and Peppers images) are downloaded from the USC-SIPI database. The popular image database BOSSbase [49] is also utilized in our testing, 1,000 images randomly selected from this database are utilized. Two indicators, i.e., the visual quality and the file size preservation are taken into account for performance evaluation. The visual quality is measured by the PSNR value in dB, and the file size preservation is measured by the increased file size in bits. For the proposed method, the parameter M controlling the choices of expansion bins is empirically taken as 6.

We first consider the impact of different values of parameter a to the embedding performance. Take those eight classical images with QF $= 70$ by embedding 10,000 bits for example. Fig. 7 shows the performance comparison with different a, where y-axes represent the average PSNR value and x-axes represent the average increased file size. Notice that, $a = 0$ means that only the block smoothness is considered while $a = 1$ means that only the band smoothness is considered. Also, for different a, the obtained results may be slightly different, but roughly the same trend. It can be observed that as a getting large, the increased file size for the marked image is getting smaller, and the least increased file size is obtained when $a = 0.3$ and $a = 0.5$. This is mainly because that as a increases, the consideration of the band smoothness TF_i increases. Correspondingly, when a grows larger than 0.5, the consideration of the band smoothness TF_i keeps increasing but the consideration of the block smoothness TB_k decreases. Therefore, when a grows larger than 0.5, the file size preservation gets worse and finally makes the largest increased file size for $a = 1$. In terms of the visual quality, the impact to the PSNR values is slight for $0 \le a \le 0.4$. However, when $a \ge 0.5$, the PSNR value decreases significantly with the increase of a. Then, considering both the PSNR and increased file size, in the following experimental comparison, we take $a = 0.3$ for the coefficient smoothness measurement based on (28).

Moreover, to illustrate the nearly optimal parameters determination process, the determined parameters $(b_1, \ldots, b_{35})$ for different images with QF $= 80$ by embedding 10,000 bits are presented in Fig. 8. Here, different parameters are demonstrated in a DCT block with different colors. Among all these parameters, the value 1 accounts for the most proportion, and it is more probable to select 1 for the smooth image like Airplane image. Meanwhile, the values 4 and 5 are rarely selected compared to the other values, since they usually account for much less in the generated histograms. As for the different histograms, in the middle frequency histograms like $\{h_3, .., h_{20}\}$, the parameters are usually selected as value 1 while values 2 and 3 are more often be taken in the other histograms. For the unmodified parameter, i.e., $+\infty$

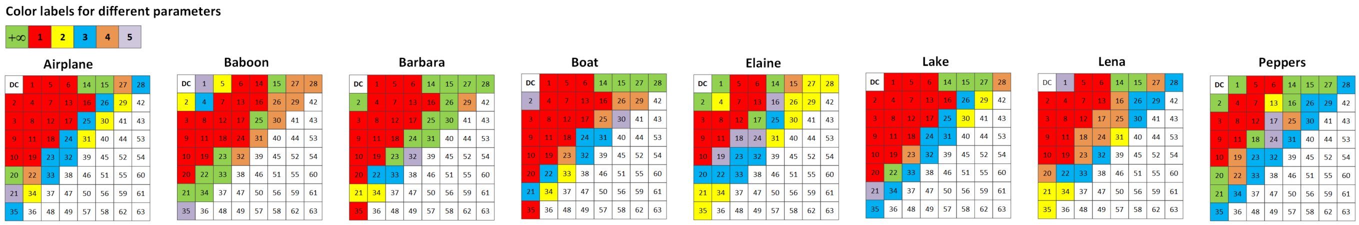

Fig. 8. Determined parameters $(b_1, ..., b_{35})$ for different images with QF $= 80$ by embedding $10,000$ bits.

labeled in green, it is more likely to be selected in some specific histograms, like h_{14} and h_{15}, which tend to be less concentrated compared with other histograms. The variation of parameters selection with different image contents also confirms the adaptive expansion bins selection of our method, and the specific comparisons with the state-of-the-art methods [40], [41], [43] are given as below.

A. Visual Quality

The proposed method is compared with [40], [41], [43] on the eight classical images with QF $= 70$, 80, and 90 firstly. The obtained PSNR values for the embedding capacity of 6,000, 9,000, 12,000, and 15,000 bits with different images for these four methods are listed in Table I, where the text in bold indicates the highest PSNR value of the four methods in the same case. It can be seen that our method achieves the highest PSNR whatever the image, quality, or capacity is, and the improvement is somewhat significant. For example, for the image Lake with QF $= 90$, our average increase of PSNR for different capacities is about 2.57, 2.77, 2.14 dB compared with [40], [41] and [43] respectively. Notice that, in some cases, Hou *et al.*'s method is nearly consistent with our method. For example, for the image Airplane with QF $= 70$ by embedding $9,000$ bits, and QF $= 90$ by embedding $9,000$ bits, the PSNR of [43] is the same as or only 0.01 dB lower than ours. However, the improvement of our method is still significant in most cases, and the average PSNR gain of the proposed method is 0.33, 0.51 and 1.09 dB for those eight images with QF $= 70$, 80, and 90 by embedding $15,000$ bits, respectively. Furthermore, to verify the superiority of our method for various images, we test the proposed method and these three methods [40], [41], [43] on database [49] with 1,000 images. The evaluations for the 1,000 images with QF $= 70$, 80, and 90 are shown in Fig. 9, where x-axes represent the embedding capacity and y-axes represent the average PSNR value. It can be observed that the proposed method outperforms [40], [41], [43] in all these cases. The reasons of our advantages are mainly about two aspects: one is that the multiple histograms are generated considering the coefficient smoothness, and another one is that the rate-distortion model is formulated for adaptive embedding.

B. File Size Preservation

The file size increase is also an important criterion for JPEG images RDH, and a better JPEG images RDH scheme should preserve this property as much as possible. The file size increases of our method and [40], [41], [43] for those eight images are given in Table II, where the text in bold indicates the lowest file size increase of the four methods in the same cases. Besides, the comparisons of the average file size increase for 1,000 images randomly selected from database [49] with QF $= 70$, 80, 90 are shown in Fig. 10.

It can be observed that the proposed method can preserve the file size better by yielding less file size increase than all these three methods for most cases, especially for high embedding capacities. This is mainly because that in our method, the coefficient selection scheme tends to select coefficients located in a smoother image region first for data embedding, in which case the selected coefficients are more likely to have with smaller zero-run-lengths. As noted in [32] and [42], embedding data into the coefficients with smaller zero-run-lengths from the blocks with fewer zero-run-value pairs can result in smaller file size increment. Besides, in our scheme, the coefficients in the high frequency are not modified at all. As mentioned above, the coefficients in high frequency are usually quantized by a large step, and many zero valued coefficients are derived to facilitate the run-length coding. Since the file size depends on the entropy encoding, in general, the more successive zero valued coefficients, the higher compression ratio can be achieved. Therefore, modify the high frequency coefficients usually result in an obvious size increase, and it is recommended to leave these high frequency coefficients unmodified for the file size preservation.

All in all, the proposed method achieves better embedding performance compared with these three methods considering both the visual quality and file size preservation.

V. Conclusion

In this paper, a novel RDH scheme for JPEG images based on MHM and rate-distortion optimization is proposed. There are three main contributions in the proposed method. First, different with previous methods, we generalize the HS-based RDH for JPEG images into a MHM embedding framework, where different histograms can be modified differently so as to make a better embedding performance. Second, a rate-distortion model which can accurately estimate the degradation on the marked JPEG image is formulated, and the expansion bins for different histograms can be adaptively determined. Third, instead of using the exhaustive search, a greedy algorithm is designed to derive the nearly optimal solution with low computation complexity. With a good file size preservation, experimental results verify that the proposed method is superior to the state-of-the-art RDH methods [40], [41],

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.2020.3027391, IEEE Transactions on Circuits and Systems for Video Technology

TABLE I

COMPARISON OF PSNR VALUES IN dB FOR DIFFERENT IMAGES WITH DIFFERENT QUALITY FACTORS FOR THE PROPOSED METHOD AND THREE STATE-OF-THE-ART METHODS [40], [41], [43]. THE TEXT IN BOLD INDICATES THE HIGHEST PSNR VALUE OF THE FOUR METHODS IN THE SAME CASE.

Image	Method	Embedding capacity (bits) with QF=70				Embedding capacity (bits) with QF=80				Embedding capacity (bits) with QF=90			
		6,000	9,000	12,000	15,000	6,000	9,000	12,000	15,000	6,000	9,000	12,000	15,000
Airplane	[40]	43.54	43.27	40.98	38.95	49.66	47.09	44.95	42.95	54.29	51.88	50.03	48.70
	[41]	44.77	42.93	40.60	39.44	48.90	46.45	44.14	42.28	53.02	50.72	48.04	47.39
	[43]	47.77	**44.68**	42.05	39.45	50.58	48.19	45.80	43.43	54.84	52.52	50.89	49.29
	Ours	**47.85**	**44.68**	**42.11**	**40.00**	**50.66**	**48.21**	**45.88**	**43.76**	**54.86**	**52.53**	**50.92**	**49.38**
Baboon	[40]	44.05	41.73	39.54	37.81	45.76	43.12	40.85	39.37	48.08	45.74	44.16	42.36
	[41]	41.82	39.81	38.43	37.48	43.54	41.49	40.24	39.06	46.09	44.38	43.12	41.87
	[43]	44.67	42.02	39.89	38.04	46.19	43.61	41.44	39.77	48.42	46.17	44.30	42.73
	Ours	**44.82**	**42.27**	**40.25**	**38.53**	**46.72**	**44.14**	**42.12**	**40.30**	**49.41**	**46.90**	**44.79**	**43.09**
Babara	[40]	45.99	42.47	39.63	37.04	48.44	46.01	43.09	40.62	52.93	50.55	48.64	46.91
	[41]	44.87	42.67	40.98	39.61	47.29	45.14	43.50	42.23	50.75	49.13	47.85	46.56
	[43]	46.59	43.39	40.96	38.69	49.52	46.62	44.02	41.97	53.73	51.17	49.14	47.11
	Ours	**46.77**	**43.72**	**41.36**	**39.63**	**49.83**	**46.92**	**44.54**	**42.48**	**53.83**	**51.58**	**49.67**	**47.77**
Boat	[40]	44.81	42.50	40.66	39.33	46.90	45.55	42.86	41.95	50.29	47.83	46.34	45.14
	[41]	44.95	42.37	40.96	39.52	46.93	44.61	43.37	42.32	50.42	48.99	47.64	46.11
	[43]	46.10	43.48	41.54	39.90	48.23	45.65	43.77	42.30	50.78	48.40	46.57	45.29
	Ours	**46.27**	**43.78**	**41.88**	**40.35**	**49.30**	**46.77**	**44.77**	**43.17**	**52.77**	**50.55**	**48.85**	**47.44**
Elaine	[40]	45.56	43.68	42.16	41.03	47.26	45.20	43.90	42.46	49.91	47.61	45.99	44.50
	[41]	46.13	44.47	43.47	42.11	48.82	46.41	45.24	44.14	50.61	49.45	47.98	46.77
	[43]	46.71	44.71	43.08	41.72	48.51	46.45	44.76	43.40	51.10	48.44	46.42	44.86
	Ours	**47.41**	**45.37**	**43.73**	**42.30**	**49.49**	**47.31**	**45.80**	**44.54**	**52.25**	**50.06**	**48.51**	**47.33**
Lake	[40]	46.31	43.52	41.29	39.47	48.41	45.84	43.94	42.20	51.06	48.12	46.12	44.66
	[41]	43.94	42.18	40.07	39.04	46.42	44.58	42.48	41.52	49.28	47.45	46.65	45.76
	[43]	46.63	43.98	41.83	39.85	49.35	46.59	44.18	42.36	51.89	48.43	46.57	44.79
	Ours	**46.80**	**44.26**	**42.14**	**40.37**	**49.52**	**46.95**	**44.84**	**42.95**	**53.64**	**50.90**	**48.74**	**46.94**
Lena	[40]	47.02	44.35	42.08	40.19	50.21	47.49	45.93	43.83	53.36	51.53	49.45	48.59
	[41]	45.97	43.97	41.83	40.38	48.19	47.10	45.16	43.44	52.73	50.48	48.92	47.71
	[43]	47.35	44.67	42.65	40.67	50.23	47.95	46.05	44.28	54.14	51.92	50.21	48.78
	Ours	**47.57**	**45.04**	**42.88**	**40.99**	**50.48**	**48.22**	**46.30**	**44.62**	**54.28**	**52.45**	**50.81**	**49.48**
Peppers	[40]	47.00	44.55	42.72	41.21	49.21	47.34	45.74	44.10	51.30	49.72	48.04	46.60
	[41]	46.94	44.22	42.72	41.44	49.28	47.21	45.59	44.20	51.76	50.26	49.20	48.05
	[43]	47.58	45.30	43.39	41.75	50.12	47.90	46.18	44.73	53.13	50.81	48.81	47.21
	Ours	**47.82**	**45.39**	**43.50**	**41.78**	**50.33**	**48.16**	**46.38**	**44.97**	**53.73**	**51.61**	**50.10**	**48.82**
Average	[40]	45.54	43.26	41.13	39.38	48.23	45.96	43.91	42.19	51.40	49.12	47.35	45.93
	[41]	44.92	42.83	41.13	39.88	47.42	45.38	43.72	42.40	50.59	48.86	47.43	46.28
	[43]	46.68	44.03	41.92	40.01	49.09	46.62	44.53	42.78	52.25	49.73	47.86	46.26
	Ours	**46.91**	**44.31**	**42.23**	**40.49**	**49.54**	**47.09**	**45.08**	**43.35**	**53.10**	**50.82**	**49.05**	**47.53**

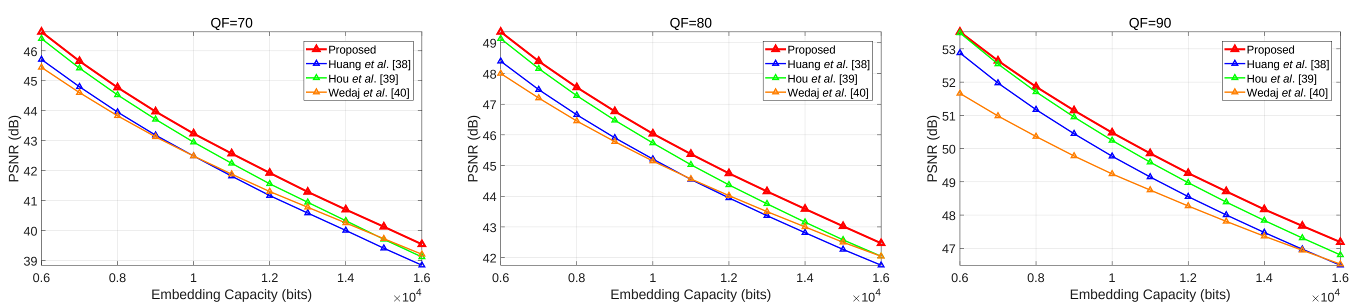

Fig. 9. Average PSNR values corresponding to different embedding capacities for the proposed method and three state-of-the-art methods [40], [41], [43].

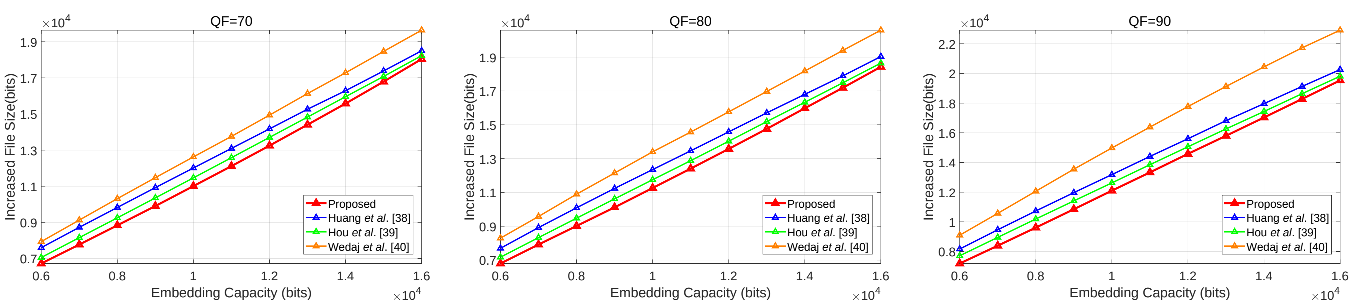

Fig. 10. Average increased file sizes corresponding to different embedding capacities for the proposed method and three state-of-the-art methods [40], [41], [43].

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.2020.3027391, IEEE Transactions on Circuits and Systems for Video Technology

TABLE II
COMPARISON OF INCREASED FILE SIZE IN BITS FOR DIFFERENT IMAGES WITH DIFFERENT QUALITY FACTORS FOR THE PROPOSED METHOD AND THREE STATE-OF-THE-ART METHODS [40], [41], [43]. THE TEXT IN BOLD INDICATES THE LOWEST FILE SIZE INCREASE OF THE FOUR METHODS IN THE SAME CASES.

Image	Method	Embedding capacity (bits) with QF=70				Embedding capacity (bits) with QF=80				Embedding capacity (bits) with QF=90			
		6,000	9,000	12,000	15,000	6,000	9,000	12,000	15,000	6,000	9,000	12,000	15,000
Airplane	[40]	6,880	10,576	13,968	17,296	6,992	10,208	13,352	16,856	8,336	12,264	15,088	18,208
	[41]	8,848	11,640	16,048	18,040	7,448	11,504	15,776	20,512	8,328	13,232	19,040	22,344
	[43]	6,848	10,416	13,896	16,912	6,648	9,336	13,160	17,000	8,416	12,016	14,592	17,944
	Ours	**6,344**	**10,248**	**13,648**	**16,432**	**6,328**	**9,288**	**12,808**	**16,576**	**7,768**	**10,992**	**14,552**	**17,824**
Baboon	[40]	7,528	10,840	15,240	19,240	7,728	11,424	15,384	18,976	8,800	12,304	16,984	21,088
	[41]	10,152	14,088	19,304	23,496	11,232	15,648	20,416	26,408	14,688	21,024	26,104	30,416
	[43]	7,264	11,072	15,112	19,272	7,552	11,256	15,216	19,088	8,728	12,480	16,792	21,160
	Ours	**6,448**	**10,008**	**14,432**	**19,040**	**6,816**	**10,504**	**15,008**	**18,688**	**8,496**	**11,968**	**15,880**	**20,200**
Babara	[40]	6,376	10,192	13,728	**18,048**	6,944	10,296	14,352	17,712	7,128	11,240	14,656	17,784
	[41]	7,504	11,720	15,376	19,616	8,544	11,968	16,680	19,936	7,320	13,528	17,888	21,304
	[43]	5,792	9,880	14,944	18,288	6,616	10,240	14,664	19,032	6,568	9,768	13,712	**16,984**
	Ours	**5,488**	**9,512**	**13,592**	19,304	**6,192**	**8,960**	**13,024**	**17,072**	**6,504**	**9,408**	**13,512**	17,624
Boat	[40]	7,432	11,008	14,584	18,416	7,864	11,336	15,104	18,744	9,240	13,656	16,952	20,696
	[41]	8,208	12,656	15,408	19,680	8,816	13,368	17,008	20,816	10,424	14,568	20,456	24,552
	[43]	6,928	10,424	13,920	17,720	7,696	11,640	15,216	18,688	10,008	14,096	17,984	21,648
	Ours	**6,760**	**9,768**	**13,856**	**17,440**	**6,768**	**9,792**	**13,552**	**17,360**	**83,36**	**12,560**	**15,632**	**20,000**
Elaine	[40]	9,952	14,480	17,920	20,800	9,464	13,864	17,432	20,904	10,120	15,256	19,384	23,488
	[41]	9,424	13,104	15,152	18,528	8,192	12,976	16,960	19,696	9,824	15,120	20,648	26,040
	[43]	9,184	13,024	15,872	18,640	8,624	12,784	16,528	19,464	10,608	14,856	19,560	23,688
	Ours	**8,232**	**12,064**	**14,552**	**17,688**	**7,792**	**11,848**	**14,944**	**17,944**	**9,256**	**13,608**	**18,480**	**22,104**
Lake	[40]	6,888	10,456	14,304	17,816	7,072	10,824	14,256	18,248	**9,048**	14,664	18,608	23,344
	[41]	9,392	12,400	17,424	20,432	94,32	13,384	17,856	21,232	11,672	17,440	22,016	26,056
	[43]	6,840	10,192	13,608	18,064	6,992	10,144	14,296	18,464	9,392	14,832	18,416	22,224
	Ours	**6,592**	**9,696**	**13,112**	**16,712**	**6,640**	**9,784**	**13,008**	**17,280**	9,408	**12,952**	**17,192**	**22,072**
Lena	[40]	8,240	11,488	14,648	17,368	8,144	11,744	14,568	18,288	9,752	13,160	16,704	19,640
	[41]	8,368	12,032	15,976	18,616	9,072	11,432	15,984	19,968	8,016	12,984	15,620	20,696
	[43]	8,024	11,368	15,112	17,408	8,200	11,328	14,944	18,200	8,488	11,456	16,144	19,432
	Ours	**7,824**	**10,600**	**14,408**	**16,480**	**7,536**	**10,496**	**13,848**	**17,424**	**7,824**	**11,040**	**14,760**	**18,008**
Peppers	[40]	8,560	12,232	15,768	18,360	9,032	12,200	14,640	18,856	9,016	12,968	16,112	20,240
	[41]	8,368	12,584	15,512	18,512	8,472	11,960	16,032	19,200	9,144	13,576	18,752	22,680
	[43]	8,760	11,680	**14,240**	18,056	8,408	11,160	14,680	**17,400**	8,840	12,272	16,720	20,976
	Ours	**8,360**	**10,968**	14,520	**17,776**	**7,640**	**10,544**	**14,336**	17,952	**7,824**	**11,424**	**15,072**	**18,648**
Average	[40]	7,732	11,409	15,020	18,418	7,905	11,487	14,886	18,573	8,930	13,189	16,811	20,561
	[41]	8,783	12,528	16,275	19,615	8,901	12,780	17,089	20,971	9,927	15,184	20,066	24,261
	[43]	7,455	11,007	14,588	18,045	7,592	10,986	14,838	18,417	8,881	12,722	16,740	20,507
	Ours	**7,006**	**10,358**	**14,015**	**17,609**	**6,964**	**10,152**	**13,816**	**17,537**	**8,177**	**11,744**	**15,636**	**19,560**

[43] in terms of capacity-distortion performance. However, one drawback the proposed method is its capacity limitation, since there are only one pair of expansion bins is used in data embedding. Then, to increase the embedding capacity, a possible direction for the future work is to extend MHM to multiple pairs of expansion bins. In addition, incorporating more reasonable smoothness measurement into our method is also a topic worthy of investigation in the future study.

REFERENCES

[1] Y.-Q. Shi, X. Li, X. Zhang, H.-T. Wu, and B. Ma, "Reversible data hiding: Advances in the past two decades," *IEEE Access*, vol. 4, pp. 3210–3237, 2016.

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

[3] W. Zhang, X. Hu, X. Li, and N. Yu, "Recursive histogram modification: Establishing equivalency between reversible data hiding and lossless data compression," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2775–2785, 2013.

[4] X. Zhang, "Reversible data hiding with optimal value transfer," *IEEE Trans. Multimedia*, vol. 15, no. 2, pp. 316–325, 2013.

[5] J. Tian, "Reversible data embedding using a difference expansion," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 8, pp. 890–896, Aug. 2003.

[6] A. M. Alattar, "Reversible watermark using the difference expansion of a generalized integer transform," *IEEE Trans. Image Process.*, vol. 13, no. 8, pp. 1147–1156, Aug. 2004.

[7] Y. Hu, H. K. Lee, K. Chen, and J. Li, "Difference expansion based reversible data hiding using two embedding directions," *IEEE Trans. Multimedia*, vol. 10, no. 8, pp. 1500–1512, 2008.

[8] S. Weng, Y. Zhao, J. S. Pan, and R. Ni, "Reversible watermarking based on invariability and adjustment on pixel pairs," *IEEE Signal Process. Lett.*, vol. 15, pp. 721–724, 2008.

[9] X. Li, W. Zhang, X. Gui, and B. Yang, "A novel reversible data hiding scheme based on two-dimensional difference-histogram modification," *IEEE Trans. Inf. Forens. Security*, vol. 8, no. 7, pp. 1091–1100, July 2013.

[10] Z. Ni, Y.-Q. Shi, N. Ansari, and W. Su, "Reversible data hiding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 3, pp. 354–362, 2006.

[11] D. M. Thodi and J. J. Rodriguez, "Expansion embedding techniques for reversible watermarking," *IEEE Trans. Image Process.*, vol. 16, no. 3, pp. 721–730, Mar. 2007.

[12] V. Sachnev, H. J. Kim, J. Nam, S. Suresh, and Y.-Q. Shi, "Reversible watermarking algorithm using sorting and prediction," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 7, pp. 989–999, Jul. 2009.

[13] X. Gao, L. An, Y. Yuan, D. Tao, and X. Li, "Lossless data embedding using generalized statistical quantity histogram," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 8, pp. 1061–1070, Aug. 2011.

[14] G. Coatrieux, W. Pan, N. Cuppens-Boulahia, F. Cuppens, and C. Roux, "Reversible watermarking based on invariant image classification and dynamic histogram shifting," *IEEE Trans. Inf. Forens. Security*, vol. 8, no. 1, pp. 111–120, Jan. 2013.

[15] C. Qin, C.-C. Chang, Y.-H. Huang, and L.-T. Liao, "An inpainting-assisted reversible steganographic scheme using histogram shifting mechanism," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 23, no. 7, pp. 1109–1118, 2013.

[16] B. Ou, X. Li, Y. Zhao, R. Ni, and Y.-Q. Shi, "Pairwise prediction-error expansion for efficient reversible data hiding," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 5010–5021, 2013.

[17] X. Li, W. Zhang, X. Gui, and B. Yang, "Efficient reversible data hiding based on multiple histograms modification," *IEEE Trans. Inf. Forens. Security*, vol. 10, no. 9, pp. 2016 – 2027, 2015.
[18] I. C. Dragoi and D. Coltuc, "Adaptive pairing reversible watermarking," *IEEE Trans. Image Process.*, vol. 25, no. 5, pp. 2420–2422, 2016.
[19] J. Wang, J. Ni, X. Zhang, and Y.-Q. Shi, "Rate and distortion optimization for reversible data hiding using multiple histogram shifting," *IEEE Trans. Cybern.*, vol. 47, no. 2, pp. 315–326, 2017.
[20] S. Kim, X. Qu, V. Sachnev, and H. Kim, "Skewed histogram shifting for reversible data hiding using a pair of extreme predictions," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 11, pp. 3236–3246, 2019.
[21] J. Wang, X. Chen, J. Ni, N. Mao, and Y. Shi, "Multiple histograms-based reversible data hiding: Framework and realization," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 8, pp. 2313–2328, 2020.
[22] S. Kim, F. Huang, and H. Kim, "Reversible data hiding in JPEG images using quantized DC," *Entropy*, vol. 21, no. 9, p. 835, 2019.
[23] D. Coltuc and J. M. Chassery, "Very fast watermarking by reversible contrast mapping," *IEEE Signal Process. Lett.*, vol. 14, no. 4, pp. 255–258, Apr. 2007.
[24] F. Peng, X. Li, and B. Yang, "Adaptive reversible data hiding scheme based on integer transform," *Signal Process.*, vol. 92, no. 1, pp. 54–62, 2012.
[25] J. Fridrich, M. Goljan, and R. Du, "Lossless data embedding for all image formats," *Ei Spie Security & Watermarking of Multimedia Contents IV*, vol. 4675, pp. 572–583, 2002.
[26] K. Wang, Z. Lu, and Y. Hu, "A high capacity lossless data hiding scheme for JPEG images," *J. Syst. Softw.*, vol. 86, no. 7, pp. 1965–1975, 2013.
[27] B. G. Mobasseri, R. J. Berger, M. P. Marcinak, and Y. J. Naikraikar, "Data embedding in JPEG bitstream by code mapping," *IEEE Trans. on Image Process.*, vol. 19, no. 4, pp. 958–966, 2010.
[28] Z. Qian and X. Zhang, "Lossless data hiding in JPEG bitstream," *J. Syst. Softw.*, vol. 85, no. 2, pp. 309–313, 2012.
[29] Y. Hu, K. Wang, and Z.-M. Lu, "An improved VLC-based lossless data hiding scheme for JPEG images," *J. Syst. Softw.*, vol. 86, no. 8, pp. 2166–2173, 2013.
[30] Z. Qian, X. Zhang, and S. Wang, "Reversible data hiding in encrypted JPEG bitstream," *IEEE Trans. Multimedia*, vol. 16, no. 5, pp. 1486–1491, Aug 2014.
[31] J. He, J. Chen, W. Luo, S. Tang, and J. Huang, "A novel high-capacity reversible data hiding scheme for encrypted JPEG bitstreams," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 12, pp. 3501–3515, 2019.
[32] Z. Qian, H. Xu, X. Luo, and X. Zhang, "New framework of reversible data hiding in encrypted JPEG bitstreams," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 2, pp. 351–362, 2019.
[33] P. Puteaux and W. Puech, "A recursive reversible data hiding in encrypted images method with a very high payload," *IEEE Trans. Multimedia*, to be published.
[34] J. Fridrich, M. Goljan, and R. Du, "Invertible authentication watermark for JPEG images," in *Proc. Int. Conf. Inf. Technol., Coding Comput*, Apr. 2001, pp. 223–227.
[35] C. Chang, C. Lin, C. Tseng, and W. Tai, "Reversible hiding in DCT-based compressed images," *Information Sciences An International Journal*, vol. 177, no. 13, pp. 2768–2786, 2007.
[36] G. Xuan, Y. Shi, Z. Ni, P. Chai, X. Cui, and X. Tong, "Reversible data hiding for JPEG images based on histogram pairs," in *Proc. Int. Conf. Image Anal. Recognit.*, 2007, pp. 715–727.
[37] H. Sakai, M. Kuribayashi, and M. Morii, "Adaptive reversible data hiding for JPEG images," in *Proc. IEEE Int. Symp. Inf. Theory Appl.*, Dec. 2008, pp. 1–6.
[38] T. Efimushkina, K. Egiazarian, and M. Gabbouj, "Rate-distortion based reversible watermarking for JPEG images with quality factors selection," in *Proc. Eur. Workshop Vis. Inf. Process.*, Jun. 2013, pp. 94–99.
[39] A. Nikolaidis, "Reversible data hiding in JPEG images utilising zero quantised coefficients," *IET Image Process.*, vol. 9, no. 7, pp. 560–568, 2015.
[40] F. Huang, X. Qu, H. J. Kim, and J. Huang, "Reversible data hiding in JPEG images," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 26, no. 9, pp. 1610–1621, 2016.
[41] F. T. Wedaj, S. Kim, H. J. Kim, and F. Huang, "Improved reversible data hiding in JPEG images based on new coefficient selection strategy," *J. Image Video Process.*, vol. 2017, no. 60, pp. 1–11, 2017.
[42] Z. Qian, S. Dai, and B. Chen, "Reversible data hiding in JPEG images using ordered embedding," *KSII Trans. Internet Inf. Syst.*, vol. 11, no. 2, pp. 945–958, 2017.
[43] D. Hou, H. Wang, W. Zhang, and N. Yu, "Reversible data hiding in JPEG image based on DCT frequency and block selection," *Signal Process.*, vol. 148, pp. 41–47, 2018.
[44] Y. Li, S. Yao, K. Yang, Y. Tan, and Q. Zhang, "A high-imperceptibility and histogram-shifting data hiding scheme for JPEG images," *IEEE Access*, vol. 7, pp. 73 573–73 582, 2019.
[45] J. He, J. Chen, and S. Tang, "Reversible data hiding in JPEG images based on negative influence models," *IEEE Trans. Inf. Forens. Security*, vol. 15, pp. 2121–2133, 2020.
[46] Z. Yin, Y. Ji, and B. Luo, "Reversible data hiding in JPEG images with multiobjective optimization," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 8, pp. 2343–2352, 2020.
[47] N. Li and F. Huang, "Reversible data hiding for JPEG images based on pairwise nonzero AC coefficient expansion," *Signal Process.*, vol. 171, p. 107476, 2020.
[48] W. Luo, J. Huang, and G. Qiu, "JPEG error analysis and its applications to digital image forensics," *IEEE Trans. Inf. Forens. Security*, vol. 5, no. 3, pp. 480–491, Sep. 2010.
[49] "Bossbase-1.01-hugo-alpha=0.4.tar.bz2." [Online], Available: http://agents.fel.cvut.cz/stegodata/, Accessed Mar. 2014.

Mengyao Xiao received his B.S. degree from Beijing Jiaotong University, China, in 2017. She is currently pursuing the Ph.D. degree in Institute of Information Science, Beijing Jiaotong University, Beijing, China. Her research interests are image processing and information hiding.

Xiaolong Li received the B.S. degree from Peking University (China), the M.S. degree from Ecole Polytechnique (France), and the Ph.D. degree in mathematics from ENS de Cachan (France), in 1999, 2002, and 2006, respectively. He worked as a postdoctoral fellow and then a researcher at Peking University in 2007-2016. He is currently a Professor with the Institute of Information Science, Beijing Jiaotong University, Beijing, China. His research interests are image processing and information hiding.

Bin Ma received the M.S. and Ph.D. degrees from Shandong University, Jinan, China, in 2005 and 2008, respectively. From 2008 to 2013, he was an associate professor with the School of Information Science, Shandong University of Political Science and Law, Jinan, China. He visited the New Jersey Institute of Technology at Newark, NJ, USA, as a visiting scholar from 2013 to 2015. He is currently a professor with the School of Cyber Security, Qilu University of Technology (Shandong Academy of Sciences), Shandong, China. He serves as an Editorial Board Member of a few journals such as the IEEE Transactions on Information Forensics and Security, the Journal of Visual Communication and Image Representation, and IEEE Signal Processing, etc. His research interests include reversible data hiding, multimedia security, and image processing. He is a member of ACM, and a member of IEEE.

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.2020.3027391, IEEE Transactions on Circuits and Systems for Video Technology

Xinpeng Zhang received the B.S. degree in computational mathematics from Jilin University, China, in 1995, and the M.E. and Ph.D. degrees in communication and information system from Shanghai University, China, in 2001 and 2004, respectively. Since 2004, he was with the faculty of the School of Communication and Information Engineering, Shanghai University, where he is currently a Professor. He is also with the faculty of the School of Computer Science, Fudan University. He was with The State University of New York at Binghamton as a Visiting Scholar from 2010 to 2011, and also with Konstanz University as an experienced Researcher, sponsored by the Alexander von Humboldt Foundation from 2011 to 2012. His research interests include multimedia security, image processing, and digital forensics. He has published over 200 papers in these areas. He was an Associate Editor of the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY from 2014 to 2017.

Yao Zhao received the Ph.D. degree from the Institute of Information Science, Beijing Jiaotong University (BJTU), Beijing, China, in 1996. He became an Associate Professor with BJTU in 1998 and became a Professor in 2001. From 2001 to 2002, he was a Senior Research Fellow with the Information and Communication Theory Group, Faculty of Information Technology and Systems, Delft University of Technology, Delft, The Netherlands. In 2015, he visited the Swiss Federal Institute of Technology, Lausanne, Switzerland (EPFL). From 2017 to 2018, he visited the University of Southern California. He is currently the Director of the Institute of Information Science, BJTU. His current research interests include image/video coding, digital watermarking and forensics, video analysis and understanding, and artificial intelligence. Dr. Zhao serves or served on the Editorial Boards of several international journals, including as an Associate Editor of the IEEE TRANSACTIONS ON CYBERNETICS, a Senior Associate Editor of the IEEE SIGNAL PROCESSING LETTERS, and an Area Editor of the Signal Processing: Image Communication. He was named as a Distinguished Young Scholar by the National Science Foundation of China in 2010, and was elected as a Chang Jiang Scholar of the Ministry of Education of China in 2013.