Reversible Data Hiding under Inconsistent Distortion Metrics

Dongdong Hou, Weiming Zhang, Yang Yang and Nenghai Yu

***Abstract*—Recursive code construction (RCC), based on the optimal transition probability matrix (OTPM), approaching the rate-distortion bound of reversible data hiding (RDH) has been proposed. Using the existing methods, OTPM can be effectively estimated only for a consistent distortion metric, i.e., if the host elements at different positions share the same distortion metric. However, in many applications, the distortion metrics are position dependent and should thus be inconsistent. Inconsistent distortion metrics can usually be quantified as a multi-distortion metric. In this paper, we first formulate the rate-distortion problem of RDH under a multi-distortion metric and subsequently propose a general framework to estimate the corresponding OTPM, with which RCC is extended to approach the rate-distortion bound of RDH under the multi-distortion metric. We apply the proposed framework to two examples of inconsistent distortion metrics: RDH in color image and reversible steganography. The experimental results show that the proposed method can efficiently improve upon the existing techniques.**

***Index Terms*—recursive code construction, rate-distortion bound, optimal transition probability matrix, inconsistent distortion metrics, multi-distortion metric, reversible steganography, reversible data hiding.**

I. Introduction

Reversible data hiding (RDH) is a special type of data hiding, whereby both the host signal and the embedded data can be restored from the marked signal without loss. This important technique is widely used in medical image [1], military image [2] and law forensics, where the original signal is so precious that it cannot be damaged. Moreover, it has been found that RDH can be quite helpful in video error-concealment coding [3], reversible image processing [4], etc.

In the past decade, multiple RDH algorithms have been proposed that can be roughly classified into three fundamental strategies: the lossless compression appending scheme [2], difference expansion [5]–[7] and histogram shift [8]. The state of the art of RDH combined such strategies with the residuals of images such as prediction errors (PEs) [9]–[15] to improve performance.

Nearly all RDH algorithms consist of two steps. The first step involves generating a host sequence with a small entropy, i.e., a sharp histogram, that can usually be achieved using PEs combined with the sorting technique [10] or a pixel selection strategy [13]. Subsequently, in the second step, users reversibly embed messages into the host sequence by modifying its histogram using methods such as difference expansion and histogram shift.

One natural difficulty of an RDH scheme concerns identifying the upper bound of the payload for the host sequence under a given distortion constraint. For an independent and identically distributed host sequence, this problem has been solved by Kalker and Willems [16], who formulated RDH as a special rate-distortion problem and obtained the rate-distortion function, i.e., the upper bound on the embedding rate under a given distortion constraint Δ, as follows:

$$\rho_{rev}(\Delta) = maximize\{H(Y)\} - H(X), \tag{1}$$

where X and Y denote the random variables of the host sequence and the marked sequence, respectively. The entropy is maximized over all transition probabilities $P_{Y|X}(y|x)$ satisfying the distortion constraint

$$\sum_{x,y} P_X(x)P_{Y|X}(y|x)d(x,y) \leq \Delta, \tag{2}$$

where $P_X(x)$ is the probability distribution of $\mathbf{X}$, and $d(x,y)$ is the defined cost of modifying x to y.

As the above implies, to evaluate the capacity of RDH under a given distortion constraint, one should first calculate the optimal transition probability matrix (OTPM) $P_{Y|X}(y|x)$ that implies the optimal modification of $\mathbf{X}$. Using OTPM, Lin *et al.* [17] proposed a coding method approaching the rate-distortion bound. By improving the recursive code construction (RCC) [16], Zhang *et al.* obtained the optimal embedding methods of RDH for binary host sequences [18], [19] and general grayscale host sequences [20] and furthermore proved that RCC will approach the rate-distortion bound as long as the adopted entropy coder reaches entropy. A coding method similar to [20] was independently presented by Zhang [21].

OTPM is essential to coding and decoding processes of the cited RCC schemes [17], [20], [21]. For certain specific distortion metrics, such as the square error distortion metric $d(x,y) = (x-y)^2$ or the L_1-Norm distortion metric $d(x,y) = |x-y|$, the corresponding OTPM has the non-crossing-edges (NCE) property [17], [22]. By relying on the NCE property, the OTPM $P_{Y|X}(y|x)$ can be analytically derived from the marginal distributions $P_X(x)$ and $P_Y(y)$. Therefore, the problem of estimating OTPM is converted to that of estimating the optimal $P_Y(y)$. To estimate $P_Y(y)$, Lin *et al.* [17] proposed the first algorithm, and Hu *et al.* [23] presented a fast algorithm. A general method for estimating OTPM without relying on the NCE property was presented in [24].

This work was supported in part by the Natural Science Foundation of China under Grant U1636201, 61572452.

D. Hou, W. Zhang, Y. Yang and N. Yu are with the School of Information Science and Technology, University of Science and Technology of China, Hefei, 230026, China. (Email: houdd@mail.ustc.edu.cn, zhangwm@ustc.edu.cn, skyyang@mail.ustc.edu.cn, ynh@ustc.edu.cn.)

Corresponding author: Weiming Zhang

A distortion metric is used to describe the cost incurred in changing the host element; we refer to position-independent metric as consistent distortion metric and to position-dependent metric as inconsistent distortion metric. For instance, assume that the host sequence consists of two pixels p_1 and p_2. When modifying the pixels by the same magnitude k, we define the cost on p_1 as $d_1(k)$ and the cost on p_2 as $d_2(k)$. If the cost functions satisfy $d_1(k) = d_2(k)$, we call the metric consistent distortion metric; otherwise, we call it inconsistent distortion metric. Although a number of algorithms have been proposed for estimating OTPM, all of them assume that the distortion metric is consistent, significantly limiting the applications of RCC schemes. In other words, for the existing RCC schemes, all elements of the host sequence, e.g., the pixels of the host image, need to share the same distortion metric. However, from the perspective of the human visual system or the security of data hiding, the changes in the smooth regions of an image are more noticeable than those in noisy regions. Therefore, the distortions caused by modifications in smooth regions and noisy regions should be different; based on this observation, position-dependent distortion metrics for steganography [25]–[28] and RDH schemes [29]–[32] have been proposed. Another example is the human eye's higher sensitivity to the green channel and lower sensitivity to the blue channel [33], implying that for RDH in color image, the distortion caused by a modification of the green channel should be defined higher than that of the blue channel. In summary, the distortion metric for each host element should be position dependent and thus inconsistent.

Several algorithms, such as [29]–[32], define inconsistent distortion metrics for RDH, which endow pixels from regions of complex texture or complex structure with lower costs. In general, such algorithms select image pixels with lower costs to carry messages, but their embedding methods are difference expansion or histogram shift without considering the optimal modification. It is usually hard to estimate the rate-distortion bound of RDH under inconsistent distortion metrics directly. However, in practical applications, inconsistent distortion metrics can be quantified as several distortion levels, i.e., a multi-distortion metric. That is to say we can well approximate inconsistent distortion metrics with a multi-distortion metric. Thus, an interesting problem concerns the estimation of the rate-distortion bound of RDH under a multi-distortion metric and solving the corresponding OTPM. When a multi-distortion metric is considered, the host sequence is classified into several subsequences according to the distortion metrics. So instead of a single histogram, multiple histograms are generated. Although RDH on multiple histograms has been studied by Hu *et al.* [34] and Li *et al.* [35], the researchers' distortion metrics are both consistent. In this paper, we first formulate the rate-distortion problem of RDH under a multi-distortion metric and subsequently present the unified framework to estimate the corresponding OTPM that includes the framework under a consistent distortion metric as a special case. We combine multiple histograms to form a compound histogram and construct a multi-distortion metric as a compound distortion metric that can convert the problem of RDH under a multi-distortion metric into that under a consistent distortion metric. Therefore, the rate-distortion problem of RDH under a multi-distortion metric can be solved using the existing methods [17], [23], [24]. The proposed framework can be used to improve RDH in various applications; we consider RDH in color image and reversible steganography as examples to show the advantages of the proposed framework.

The rest of this paper is organized as follows. Section II briefly introduces the existing methods for optimal RDH under a consistent distortion metric. In Section III, we formulate the rate-distortion problem of RDH under a multi-distortion metric and describe a solution to this problem. In Section IV, two applications, RDH in color image and reversible steganography, are presented to demonstrate the power of the proposed framework. Finally, this paper is concluded with the discussion in Section VI.

II. Existing Methods

Throughout this paper, matrices and vectors are shown in bold. The sender embeds L bits of a message into the host sequence $\mathbf{X} = (x_1, \cdots, x_N)$ by slightly modifying its elements to produce a marked sequence $\mathbf{Y} = (y_1, \cdots, y_N)$. We denote the embedding rate $R = L/N$, where N is the length of the host sequence. Schemes are usually constructed to minimize the average distortion between $\mathbf{X}$ and $\mathbf{Y}$ for a given embedding rate R. The cost of changing x to y is defined as $d(x, y)$, which could be the square error distortion metric $d_s(x, y) = (x - y)^2$, the L_1-Norm distortion metric $d_1(x, y) = |x - y|$, or a specific distortion metric defined by the user. We use X and Y to denote the random variables of the host sequence and the marked sequence, respectively; the probability distribution of the host sequence $P_X(x)$ can be estimated by the histogram of the host sequence $\mathbf{X}$. We denote the entropy of $\mathbf{X}$ by $H(X)$ and the conditional entropy of $\mathbf{Y}$ given $\mathbf{X}$ by $H(Y|X)$.

The optimization problem in Eq. (1) is for a distortion-limited sender. In practice, we usually consider a payload-limited sender, as shown in Eq. (3), that minimizes the average distortion for a given embedding rate R.

$$\begin{array}{ll} \text{minimize} & \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} P_X(x) P_{Y|X}(y|x) d(x, y) \\ \text{subject to} & H(Y) = R + H(X) \end{array} \tag{3}$$

As $x \in \mathcal{X} = \{0, 1, \cdots, m-1\}$, $y \in \mathcal{Y} = \{0, 1, \cdots, n-1\}$, and $\mathcal{X}$ and $\mathcal{Y}$ are both finite alphabet sets, the distortion metric $d(x, y)$ can be described with a distortion matrix such as

$$\mathbf{D} = \begin{bmatrix} d(0,0) & d(0,1) & ... & d(0,n-1) \\ d(1,0) & d(1,1) & ... & d(1,n-1) \\ \vdots & \vdots & \vdots & \vdots \\ d(m-1,0) & d(m-1,1) & ... & d(m-1,n-1) \end{bmatrix}, \tag{4}$$

where $d(x, y)$ in Eq. (4) has the same function form for all $x = 0, , m-1$ and $y = 0, , n-1$.

For several specific distortion metrics, such as the squared error distortion metric and the L_1-Norm distortion metric, it has been proven in [17], [22] that the corresponding OTPM

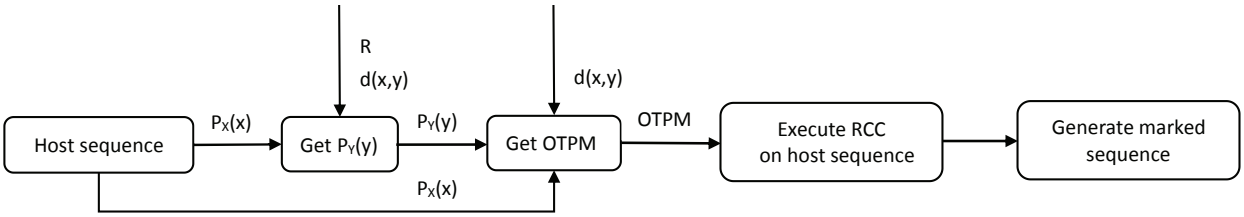


Fig. 1: The primary steps of designing RDH schemes approaching the rate-distortion bound.

$P_{Y|X}(y|x)$ can be analytically expressed by the host distribution $P_X(x)$ and the marginal distribution $P_Y(y)$ such that

$$P_{Y|X}(y|x) = \max\{0, \min\{F_X(x), F_Y(y)\} - \max\{F_X(x-1), F_Y(y-1)\}\}, \quad (5)$$

where $F_X(x)$ and $F_Y(y)$ are cumulative probability distributions of $\mathbf{X}$ and $\mathbf{Y}$, defined by $F_X(x) = \sum_{i=0}^{x} P_X(i)$, where $x = 0, \cdots, m-1$, and $F_Y(y) = \sum_{i=0}^{y} P_Y(i)$, where $y = 0, \cdots, n-1$. Note that $F_X(-1) = F_Y(-1) = 0$, $F_X(m-1) = F_Y(n-1) = 1$, the function $min\{a, b\}$ returns the minimum value of a and b, and the function $max\{a, b\}$ returns the maximum value of a and b.

Lin *et al*. [17] proposed a backward and forward iterative algorithm for estimating the optimal $P_Y(y)$, while Hu *et al*. [23] proposed a fast algorithm for estimating the optimal $P_Y(y)$ based on the Lagrangian duality. The general framework for estimating the OTPM $P_{Y|X}(y|x)$ from $P_X(x)$ and $P_Y(y)$ under any consistent distortion metrics was proposed by Zhang *et al*. [24]. According to the OTPM, we can reversibly embed messages and minimize the average distortion with an RCC scheme such as RHM (recursive histogram modification) [20]. Fig. 1 describes the primary steps of designing an RDH scheme approaching the rate-distortion bound.

Note that all the above-mentioned methods [17], [23], [24] for estimating OTPM require that the distortion metric is consistent and can be expressed as a single distortion matrix of Eq. (4). However, as discussed in the Introduction, distortion metrics in many applications are usually inconsistent, and inconsistent distortion metrics can be quantified as a multi-distortion metric. Therefore, it is important to estimate OTPM and implement the optimal embedding for RDH under a multi-distortion metric.

III. OTPM OF RDH UNDER A MULTI-DISTORTION METRIC

A. Rate-distortion bound under a multi-distortion metric

In most cases, the distortion caused by modifying each host element should be associated with its position and neighboring elements. Assuming that the distortion metric for the jth element x_j is $d_j(x, y)$, where $x_j \in \mathcal{X} = (x_1, x_2, \cdots, x_N)$ and $1 \le j \le N$, the corresponding distortion matrix $\mathbf{D}_j$ is

$$\mathbf{D}_j = \begin{bmatrix} d_j(0,0) & d_j(0,1) & ... & d_j(0,n-1) \\ d_j(1,0) & d_j(1,1) & ... & d_j(1,n-1) \\ \vdots & \vdots & \vdots & \vdots \\ d_j(m-1,0) & d_j(m-1,1) & ... & d_j(m-1,n-1) \end{bmatrix}. \quad (6)$$

There will be at most N inconsistent distortion metrics, denoted by $(d_1(x, y), d_2(x, y), \cdots, d_N(x, y))$. If N distortion metrics are identical, such a model will be one with a consistent distortion metric.

In practical applications, as many elements of the host sequence will share the same or similar distortion metrics, we can cluster such distortion metrics in K ($K \le N$) classes. Accordingly, the host sequence $\mathbf{X}$ 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 $d_i(x, y)$, which can also be described with the single distortion matrix $\mathbf{D}_i$, where $1 \le i \le K$.

Now, the host sequence $\mathbf{X}$ contains K subsequences and thus contains K histograms, with $P_{X_i}(x)$ representing the probability distribution of $\mathbf{x}_i$, where $1 \le i \le K$. However, the shapes and distortion metrics of such histograms both vary. For a fixed payload R, different histograms perform differently when chosen for embedding. As a result, a payload allocation problem arises naturally.

For each subsequence $\mathbf{x}_i$, if we obtain its allocated embedding rate denoted by R_i (relative to $\mathbf{x}_i$), we can calculate the corresponding sub-OTPM denoted as $P_{Y_i|X_i}(y|x)$ using the existing methods for a consistent distortion metric and subsequently optimally modify its histogram to obtain the corresponding marked subsequence denoted by $\mathbf{y}_i$. The total distortion caused by embedding the payload R_i into $\mathbf{x}_i$ (denoted by J_i) is

$$J_i = N_i \sum_{x,y} P_{X_i}(x) P_{Y_i|X_i}(y|x) d_i(x, y). \quad (7)$$

As for the entire payload R, the corresponding payload carried by the subsequence $\mathbf{x}_i$ is $\frac{N_i \times R_i}{N}$. Of course, the sum of payloads of K subsequences equals the total payload R, i.e.,

$$R = \frac{\sum_{i=1}^{K} N_i \times R_i}{N}. \quad (8)$$

Based on the above discussion, given a payload R, the problem of reasonably distributing the total payload R among K subsequences to minimize the average embedding distortion can be formulated as

Problem I

$$\begin{aligned} &\text{minimize} && \frac{\sum_{i=1}^{K} N_i \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} P_{X_i}(x) P_{Y_i|X_i}(y|x) d_i(x,y)}{N} \\ &\text{subject to} && \frac{\sum_{i=1}^{K} N_i \times H(Y_i)}{N} = R + \frac{\sum_{i=1}^{K} N_i \times H(X_i)}{N} \end{aligned}. \quad (9)$$

Eq. (9) provides the rate-distortion bound of RDH under a multi-distortion metric, matching that of Eq. (3) if $K = 1$.

To minimize the average distortion, each optimal allocated payload R_i, i.e., the sub-OTPM $P_{Y_i|X_i}(y|x)$, is needed for $i = 1, 2, ..., K$, which can be used to optimally modify K subsequences and embed the corresponding messages into each host subsequence.

B. Compound histogram and compound distortion metric

The host sequence $\mathbf{X}$ is divided into K subsequences according to their distortion metrics, while each subsequence has its own distortion metric and probability distribution. The difficulty lies in obtaining the optimal payload for each subsequence, i.e., obtaining the respective sub-OTPMs. To solve this problem, we combine K subsequences to form a compound sequence denoted by $\mathbf{X}_c$ and then design a compound distortion metric denoted by $d_c(x, y)$ for $\mathbf{X}_c$ that can be used to solve the optimization problem of Eq. (9) with existing methods [17], [23], [24] for a consistent distortion metric.

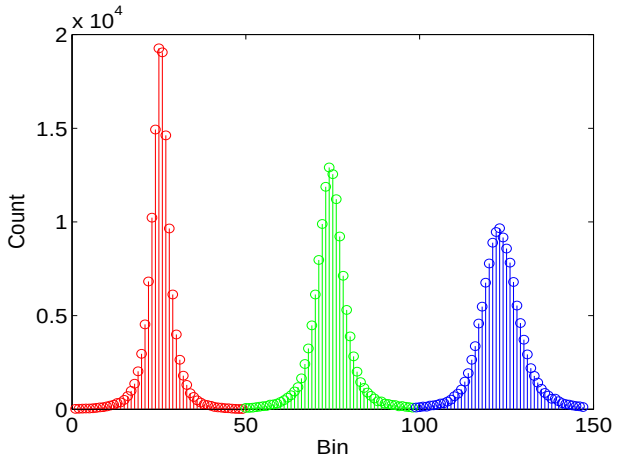


Fig. 2: Compound histogram.

Assume that the histogram of the subsequence $\mathbf{x}_i$ is denoted by $\mathbf{H}_i$; we assign it an offset value ox_i $(i = 1, 2, \cdots, K)$. By translating each sub-histogram $\mathbf{H}_i$ by the corresponding offset ox_i along the x-axis, we combine K sub-histograms $\mathbf{H}_i$ into a large compound histogram $\mathbf{H}_c$. As depicted in Fig. 2, three sub-histograms, shown in red, green and blue, are combined to form a single compound histogram. There is no need to record the offset value ox_i, as it is uniquely determined. Specifically, ox_i is selected to guarantee that there is neither overlap nor vacant space between any two adjacent sub-histograms of the compound histogram $\mathbf{H}_c$. In the proposed method, $ox_i = (i-1)m$ usually. Each sub-histogram $\mathbf{H}_i$, i.e., the subsequence $\mathbf{x}_i$, is translated by the corresponding offset ox_i along the x-axis. The translated $\mathbf{x}_i$ is denoted by $\mathbf{x}_i^t$, with its elements being in the range $\{(i-1)m, (i-1)m+1, \cdots, im-1\}$. Subsequently, K translated subsequences are combined to generate a compound sequence $\mathbf{X}_c$, as Eq. (10) shows.

$$\mathbf{X}_c = [\mathbf{x}_1^t \quad \mathbf{x}_2^t \quad ... \quad \mathbf{x}_K^t] \tag{10}$$

By normalizing the compound histogram $\mathbf{H}_c$, the compound host probability distribution $P_{X_c}(x)$ is obtained. Note that after the translation, the range of x will be enlarged by a factor of K, and $x \in \mathcal{X}^c = \{0, 1, 2, \cdots, km-1\}$. We denote the compound host probability distribution $P_{X_c}(x)$ in the vector as $\mathbf{P}_{X_c} = [\mathbf{P}_{X_1^t}, \cdots, \mathbf{P}_{X_K^t}]$, where

$$\mathbf{P}_{X_i^t} = [P_{X_c}((i-1)m), \cdots, P_{X_c}(im-1)], 1 \le i \le K. \tag{11}$$

As the probability distribution $\mathbf{P}_{X_c}$ results from normalizing the compound histogram $\mathbf{H}_c$, we obtain

$$\mathbf{P}_{X_i^t} = \frac{N_i}{N}\mathbf{P}_{X_i}, 1 \le i \le K. \tag{12}$$

Accordingly, the range of the marked sequence y is also enlarged by the factor of K, and $y \in \mathcal{Y}^c = \{0, 1, 2, \cdots, kn-1\}$. We define the compound probability distribution of the marked sequence as $\mathbf{P}_{Y_c} = [\mathbf{P}_{Y_1^t}, \cdots, \mathbf{P}_{Y_K^t}]$, where

$$\mathbf{P}_{Y_i^t} = [P_{Y_c}((i-1)n), \cdots, P_{Y_c}(in-1)], 1 \le i \le K. \tag{13}$$

We need to optimally modify the host subsequences $\mathbf{x}_i^t$ to generate the corresponding marked subsequences $\mathbf{y}_i^t$ for $i = 1, 2, ..., K$. Note that $\mathbf{y}_i^t$ must be generated only from $\mathbf{x}_i^t$; otherwise, the modification will be meaningless. To avoid modifying the elements of $\mathbf{x}_i^t$ to generate the elements of $\mathbf{y}_j^t$ when $i \neq j$, we define the cost of such a modification as infinite. The distortion matrix for the compound host sequence under a multi-distortion metric is shown in Eq. (14), and $d_c(x, y)$ represents the element in xth row and yth column of $\mathbf{D}_c$.

$$\mathbf{D}_c = \begin{bmatrix} \mathbf{D}_1 & \infty & ... & \infty \\ \infty & \mathbf{D}_2 & ... & \infty \\ \vdots & \vdots & \vdots & \vdots \\ \infty & \infty & ... & \mathbf{D}_K \end{bmatrix}, \tag{14}$$

where $\mathbf{D}_i$ is shown in Eq. (15).

By constructing the compound histogram and the compound distortion metric, we convert the problem of RDH under a multi-distortion metric into the problem under a consistent distortion metric. The constructed compound distortion metric can also be represented in a single distortion matrix $\mathbf{D}_c$. In fact, we convert the optimization Problem I of Eq. (9) into Problem II of Eq. (16).

Problem II

$$\begin{aligned} &\text{minimize} && \textstyle\sum_{x=0}^{Km-1}\sum_{y=0}^{Kn-1} P_{X_c}(x)P_{Y_c|X_c}(y|x)d_c(x,y) \\ &\text{subject to} && H(Y_c) = R + H(X_c) \end{aligned} \tag{16}$$

The problem in Eq. (16) has the same form as that in Eq. (3); hence, we can estimate the optimal distribution $P_{Y_C}(y)$ and the OTPM $P_{Y_c|X_c}(y|x)$ using methods of [17], [23], [24]. By introducing the infinite distortion, in the optimal modification that causes the minimal distortion, the elements of $\mathbf{x}_i^t$ will be not modified to the elements of $\mathbf{y}_j^t$ for $i \neq j$. Therefore, the corresponding sub-OTPM among $\mathbf{x}_i^t$ and $\mathbf{y}_j^t$ is zero, and the compound OTPM $P_{Y_c|X_c}(y|x)$ has the following form:

$$\mathbf{P}_{Y_c|X_c} = \begin{bmatrix} \mathbf{P}_{Y_1^t|X_1^t} & 0 & ... & 0 \\ 0 & \mathbf{P}_{Y_2^t|X_2^t} & ... & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & ... & \mathbf{P}_{Y_K^t|X_K^t} \end{bmatrix}, \tag{17}$$

$$\mathbf{D}_i = \begin{bmatrix} d_i((i-1)m,(i-1)n) & d_i((i-1)m,(i-1)n+1) & ... & d_i((i-1)m,in-1) \\ d_i((i-1)m+1,(i-1)n) & d_i((i-1)m+1,(i-1)n+1) & ... & d_i((i-1)m+m-1,in-1) \\ \vdots & \vdots & \vdots & \vdots \\ d_i(im-1,(i-1)n) & d_i(im-1,(i-1)n+1) & ... & d_i(im-1,in-1) \end{bmatrix} \tag{15}$$

$$\mathbf{P}_{Y_i^t|X_i^t} = \begin{bmatrix} P_{Y_c|X_c}((i-1)n|(i-1)m) & P_{Y_c|X_c}((i-1)n+1|(i-1)m) & ... & P_{Y_c|X_c}(in-1|(i-1)m) \\ P_{Y_c|X_c}((i-1)n|(i-1)m+1) & P_{Y_c|X_c}((i-1)n+1|(i-1)m+1) & ... & P_{Y_c|X_c}((in-1|(i-1)m+1) \\ \vdots & \vdots & \vdots & \vdots \\ P_{Y_c|X_c}((i-1)n|im-1) & P_{Y_c|X_c}((i-1)n+1|im-1) & ... & P_{Y_c|X_c}(in-1|im-1) \end{bmatrix} \tag{18}$$

where $\mathbf{P}_{Y_i^t|X_i^t}$ is shown in Eq. (18), and $P_{Y_c|X_c}(y|x)$ represents the element in the xth row and yth column of $\mathbf{P}_{Y_c|X_c}$.

After obtaining the optimal distribution $\mathbf{P}_{Y_C}$ and the OTPM $\mathbf{P}_{Y_c|X_c}$, the allocated payload for each subsequence is determined, according to which we can optimally modify the respective K subsequences using RCC schemes. To this end, RCC needs to be performed K times. Indeed, we can also directly apply RCC to the compound host sequence $\mathbf{X}_c$ to embed messages and generate the marked sequence $\mathbf{Y}_c$ according to $\mathbf{P}_{Y_c|X_c}$. After obtaining $\mathbf{Y}_c$, we need to retranslate each translated marked subsequence $\mathbf{y}_i^t$ for $i = 1, 2, ..., K$ by its corresponding offset oy_i (similarly, $oy_i = (i-1)n$ usually) along the x-axis to create the ultimate marked sequence $\mathbf{Y}$.

The above analyses show that we can obtain the optimal solution of Problem II using the existing methods [17], [23], [24]; the theoretical and experimental proofs that the optimal solution of Problem II can yield the optimal solution of Problem I are provided in the Appendix.

C. Embedding and extraction

Thus far, we have proposed a general framework for calculating the OTPM of RDH under a multi-distortion metric, enabling us to embed messages using the optimal coding methods such as RHM [20]. The proposed framework can be implemented in various forms; a typical implementation of a design of an RDH algorithm approaching the rate-distortion bound of RDH under a multi-distortion metric proceeds as follows.

Embedding

1) Reserve several elements of the host signal to embed the auxiliary information, and generate the host sequence $\mathbf{X}$ from the remaining elements. The auxiliary information will be embedded into the LSBs (least significant bits) of the reserved elements, with the LSBs of such reserved elements embedded into $\mathbf{X}$ as part of the payload.
2) Divide the host sequence $\mathbf{X}$ into K classes according to their distortion metrics, and obtain the subsequence $\mathbf{x}_i$ for $i = 1, 2, ..., K$.
3) Translate each subsequence $\mathbf{x}_i$ by the corresponding offset ox_i to obtain $\mathbf{x}_i^t$, and normalize the combined compound host histogram $\mathbf{H}_c$ to generate the compound host probability distribution $\mathbf{P}_{X_c}$.
4) Construct the compound distortion matrix $\mathbf{D}_c$ as Eq. (14).
5) With $\mathbf{P}_{X_c}$, $\mathbf{D}_c$ and the embedding rate R as parameters, calculate the optimal compound distribution of the marked sequence $\mathbf{P}_{Y_c}$ and the OTPM $\mathbf{P}_{Y_c|X_c}$.
6) According to $\mathbf{P}_{Y_c|X_c}$, embed messages into $\mathbf{X}_c$ with RHM [20] to generate the compound marked sequence $\mathbf{Y}_c$.
7) Retranslate each marked subsequence $\mathbf{y}_i^t$ for $i = 1, 2, ..., K$ by the corresponding offset oy_i to create the ultimate marked sequence $\mathbf{Y}$.
8) Embed the auxiliary information into the LSBs of reserved elements, primarily the location map of overflow/underflow pixels, the host histogram $\mathbf{P}_{X_c}$, the embedding rate R, and the number of classes K.

Extraction

1) Extract auxiliary information from the reserved elements, including the location map of overflow/underflow pixels, the host histogram $\mathbf{P}_{X_c}$, the embedding rate R, and the number of classes K.
2) Divide the marked sequence $\mathbf{Y}$ into K classes according to the distortion metrics, and obtain the marked subsequence $\mathbf{y}_i$ for $i = 1, 2, ..., K$.
3) Translate each marked subsequence $\mathbf{y}_i$ for $i = 1, 2, ..., K$ by the corresponding offset oy_i, and combine the results to generate the compound marked sequence $\mathbf{Y}_c$.
4) Using $\mathbf{P}_{X_c}$, embedding rate R and the constructed $\mathbf{D}_c$ as parameters, calculate OTPM $\mathbf{P}_{Y_c|X_c}$. Using $\mathbf{P}_{Y_c|X_c}$, decode the compound marked sequence $\mathbf{Y}_c$ to extract the embedded messages and restore the compound host sequence $\mathbf{X}_c$.
5) Retranslate each host subsequence $\mathbf{x}_i^t$ for $i = 1, 2, ..., K$ by the corresponding offset ox_i to reconstruct the host sequence $\mathbf{X}$. Finally, reconstruct the reserved elements using the extracted LSBs.

IV. Applications to inconsistent distortion metrics

At the beginning of this section, we first declared certain settings for RHM [20] that can approach the rate-distortion bound of RDH according to OTPM. As certain auxiliary information guaranteeing reversibility, such as the host histogram and the location map of overflow/underflow pixels, is needed for RHM (see [20]), we should omit extra payload to carry such auxiliary information. Usually, the host histogram

accounts for most of the auxiliary information. In [24], the differential pulse-code modulation encoder is used to compress the host histogram; however, we think that the host histogram can be compressed more efficiently with the help of the histogram of the marked sequence. Indeed, after modifying the host sequence to generate the marked sequence, the marked histogram will be very similar to the host histogram, especially when the payload is low. The marked histogram can be reconstructed from the marked sequence; we can further restore the host histogram from the marked histogram by calculating and recording the difference between each bin of the former and the corresponding bin of the latter. Thus, to record the host histogram, we only need to compress and record such bin differences, allowing efficient compression of the host histogram.

In the experiments, assuming the length of raw messages to be L bits, we allocate $0.03L$ bits for auxiliary messages to record the location map of overflow/underflow pixels. Then, the embedding rate for calculating $\mathbf{P}_{Y_C}$ and $\mathbf{P}_{Y_c|X_c}$ is $R = \frac{1.03L+L_{para}}{N_h}$, where N_h is the length of the host sequence, and L_{para} denotes the amount of information for recording parameters, including the host histogram, the embedding rate and the number of classes. Therefore, the number of reserved elements is usually initialized at $0.03L + L_{para}$. If it is insufficient, we can further increase the number of reserved elements.

Using the embedding rate R and the compound distortion matrix $\mathbf{D}_c$, we apply the fast algorithm by Hu *et al*. [23] to calculate the optimal probability distribution of the compound marked sequence, i.e., $\mathbf{P}_{Y_C}$. The real embedding rate corresponding to $\mathbf{P}_{Y_C}$ is $(H(Y_c) - H(X_c))$ and may be close to R, but not exactly equal to R because of the numerical precision. To solve this problem in RCC schemes, a series of embedding rates denoted by $Rtest$s are tested until

$$0 \leq H(Y_c) - H(X_c) - R \leq 0.005. \tag{19}$$

The details of the process used to calculate $\mathbf{P}_{Y_c}$ and guarantee the raw payload are shown in Algorithm 1.

Algorithm 1 Guaranteeing the Raw Payload

Input: The compound host probability distribution $\mathbf{P}_{X_c}$ and the length of raw messages L.

Output: The compound marked probability distributions, i.e., $\mathbf{P}_{Y_c}$.

$R = \frac{1.03L+L_{para}}{N_h}$;
$tag = true$, $iteration = 0$, $Rtest = R$;
while tag **do**
  $iteration = iteration + 1$;
  if $iteration > 6$ **then**
    $tag = false$;
  end
  Calculate $\mathbf{P}_{Y_C}$ using the fast algorithm by Hu *et al*. [23] with $\mathbf{P}_{X_c}$, $\mathbf{D}_c$ and $Rtest$ as parameters;
  if $0 \leq H(Y_c) - H(X_c) - R \leq 0.005$ **then**
    $tag = false$;
  else
    $Rtest = Rtest\frac{R}{H(Y_c)-H(X_c)}$;
  end
end
return $\mathbf{P}_{Y_C}$;

A. Reversible data hiding in color image

RDH algorithms in grayscale image have been well-established, in contrast to those in color image, despite the greater popularity of the latter. Among color RDH algorithms [36]–[38], the method of Yao *et al*. [38] is state of the art. In the scheme of Yao *et al*., pixels of each color channel are divided into two sets labeled dark and white, as shown in Fig. 3. Therefore, red, green and blue channels are divided into 6 sets, denoted by $\mathbf{R}_d$, $\mathbf{R}_w$, $\mathbf{G}_d$, $\mathbf{G}_w$, $\mathbf{B}_d$, and $\mathbf{B}_w$, where $\mathbf{C}_d$ and $\mathbf{C}_w$ represent the sets of scales in channel C from the dark region and the white region, with C representing the red, green or blue channel. Before embedding, the payload is allocated adaptively to the 6 divided sets to reduce the embedding distortion; subsequently, messages are embedded into the 6 sets one by one. When a given set is selected for modification, the other 5 sets are used as references to generate a sharp PE histogram through the guided filtering predictor

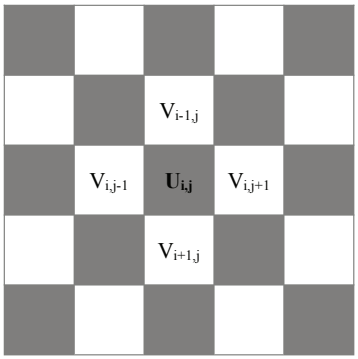
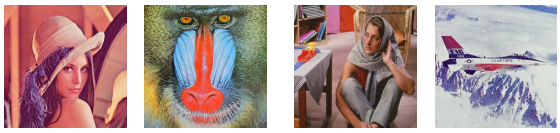


Fig. 3: A checkerboard pattern.



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

Fig. 4: Tested color images of size 510×510: (a) Lena, (b) Baboon, (c) Barbara and (d) Airplane.

[39]. Studies by [39] and [38] provide the details of the guided filtering predictor.

The existing color RDH methods [36]–[38] embed messages in a color image primarily by exploring the correlations among the three color channels to achieve a high peak signal-to-noise-ratio (PSNR) for the color marked image, implying that for these methods, modifications of the three channels have the same impact on image quality. However, the sensitivity of human eye to colors varies with color. The formula in Eq. (20) from Rec. 601 [33] is widely used for converting a color image to grayscale, with the weights in Eq. (20) representing the

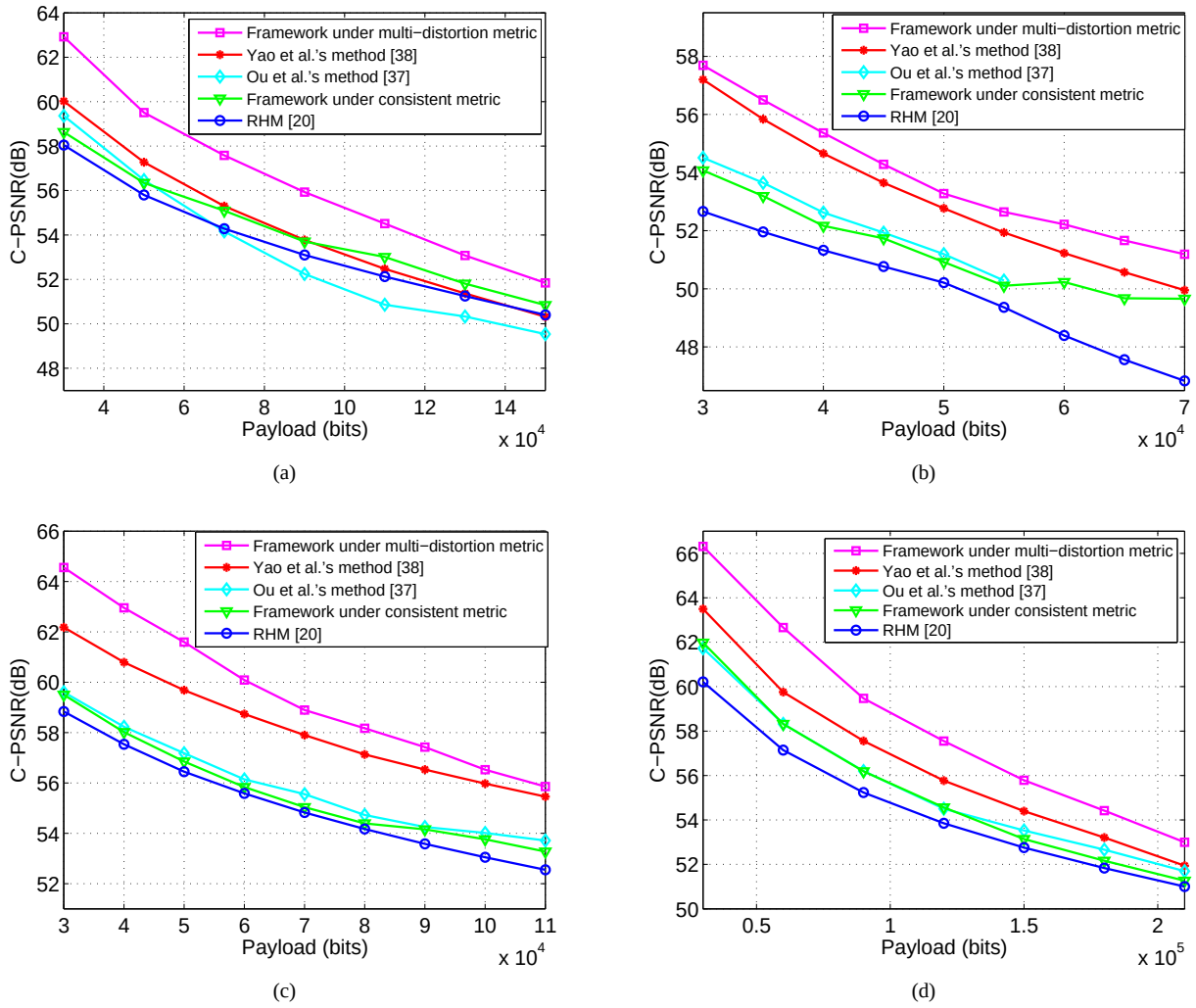


Fig. 5: Comparisons of C-PSNR: (a) Lena, (b) Baboon, (c) Barbara and (d) Airplane.

relative luminance perceptions of typical trichromat humans to light of the precise additive primary colors.

$$f = 0.299r + 0.587g + 0.114b, \tag{20}$$

where r, g and b are the levels of the red, green and blue channel, respectively, and f is the generated grayscale level. Eq. (20) shows that human eye is more sensitive to green and less so to blue. According to Eq. (20), a reasonable quality assessment of color marked image, denoted by C-PSNR, is computed as

$$C - PSNR = 10log_{10}(\frac{255^2}{MSE_C}), \tag{21}$$

where

$$MSE_C = 0.299MSE_R + 0.587MSE_G + 0.114MSE_B, \tag{22}$$

and MSE_R, MSE_G and MSE_B are the mean square errors from the red, green and blue channels, respectively.

Based on the above quality assessment, a reasonable multi-distortion metric for three host subsequences for the red, green and blue channels can be defined as

$$\begin{cases} d_R(x,y) = 0.299(x-y)^2, & for\ red\ channel \\ d_G(x,y) = 0.587(x-y)^2, & for\ green\ channel \\ d_B(x,y) = 0.114(x-y)^2, & for\ blue\ channel \end{cases}. \tag{23}$$

As in most RDH methods, PEs of pixels are generated as a host sequence to carry messages. To use the correlations among the three color channels, we also adopt the guided filter [39] used in the method of Yao *et al.* to generate host PEs. Similar to the method of Yao *et al.*, the color host image is divided into 6 sets. Considering embedding in dark regions as an example, the PEs of sets $\mathbf{R}_d$, $\mathbf{G}_d$ and $\mathbf{B}_d$ are first calculated by a guided filter with the original host image as a reference using the formulas

$$\begin{cases} \mathbf{PER}_d^0 = GuideFilter(\mathbf{R}_d, \{\mathbf{R}_w, \mathbf{G}_d, \mathbf{G}_w, \mathbf{B}_d, \mathbf{B}_w\}) \\ \mathbf{PEG}_d^0 = GuideFilter(\mathbf{G}_d, \{\mathbf{R}_d, \mathbf{R}_w, \mathbf{G}_w, \mathbf{B}_d, \mathbf{B}_w\}) \\ \mathbf{PEB}_d^0 = GuideFilter(\mathbf{B}_d, \{\mathbf{R}_d, \mathbf{R}_w, \mathbf{G}_d, \mathbf{G}_w, \mathbf{B}_w\}) \end{cases}. \tag{24}$$

For the three sets of PEs, $\mathbf{PER}_d^0$, $\mathbf{PEG}_d^0$ and $\mathbf{PEB}_d^0$ from

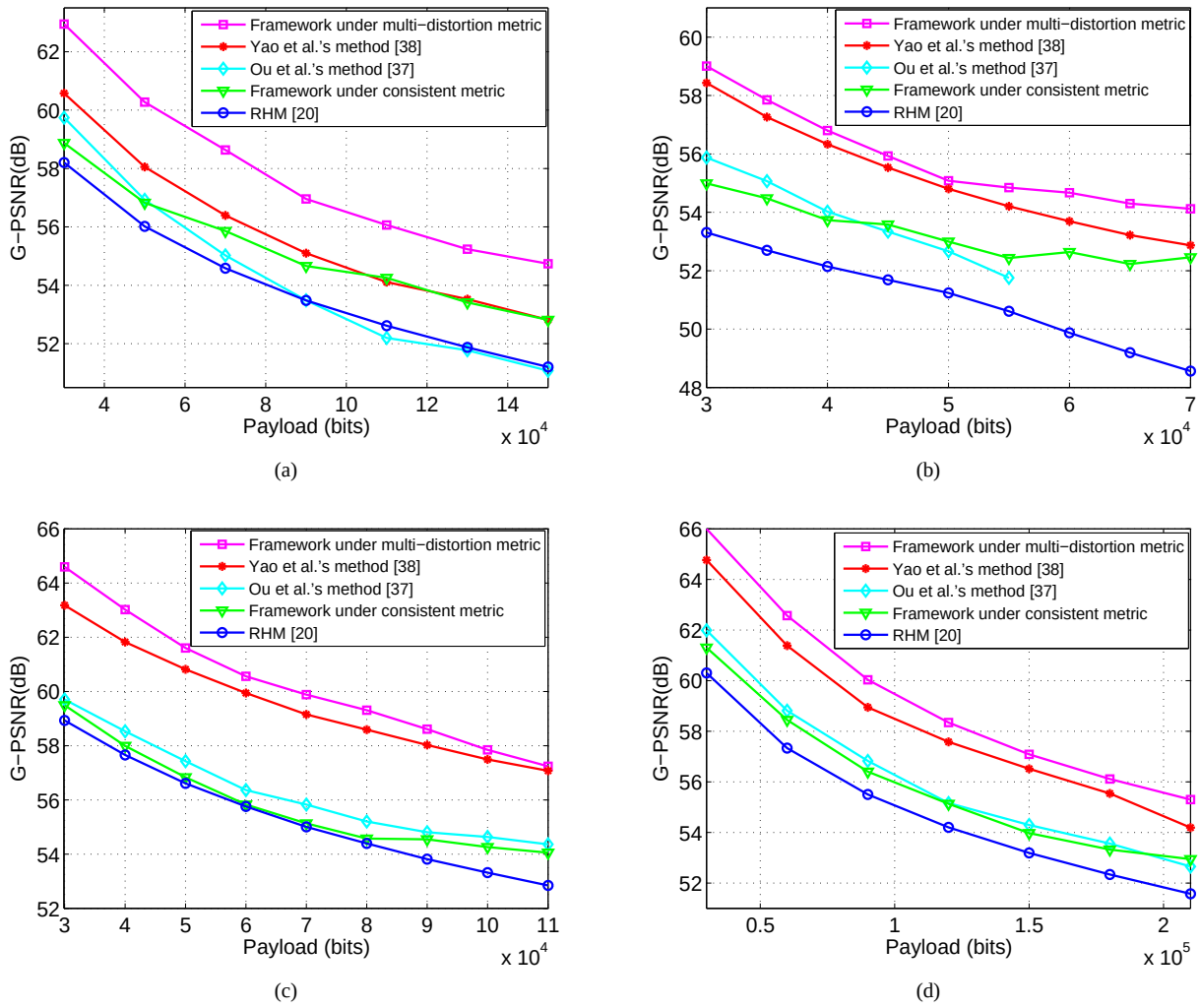


Fig. 6: Comparisons of G-PSNR: (a) Lena, (b) Baboon, (c) Barbara and (d) Airplane.

the red, green and blue channels, we define the multi-distortion metric following Eq. (23). After the payload is given, we apply our framework to estimate the optional allocated payloads for $\mathbf{R}_d$, $\mathbf{G}_d$ and $\mathbf{B}_d$.

After obtaining the allocated messages for pixel sets $\mathbf{R}_d$, $\mathbf{G}_d$ and $\mathbf{B}_d$, we use RHM [20] to embed the allocated messages into $\mathbf{PER}_d^0$ first and correspondingly modify $\mathbf{R}_d$ to $\mathbf{R}'_d$. Subsequently, we calculate the PEs of $\mathbf{G}_d$, denoted by $\mathbf{PEG}_d$:

$$\mathbf{PEG}_d = GuideFilter(\mathbf{G}_d, \{\mathbf{R}'_d, \mathbf{R}_w, \mathbf{G}_w, \mathbf{B}_d, \mathbf{B}_w\}). \quad (25)$$

After finishing embedding for $\mathbf{PEG}_d$ according to its allocated messages, $\mathbf{G}_d$ is modified to $\mathbf{G}'_d$. Finally, the PEs of $\mathbf{B}_d$, denoted by $\mathbf{PEB}_d$, are generated as follows:

$$\mathbf{PEB}_d = GuideFilter(\mathbf{B}_d, \{\mathbf{R}'_d, \mathbf{R}_w, \mathbf{G}'_d, \mathbf{G}_w, \mathbf{B}_w\}), \quad (26)$$

and $\mathbf{B}_d$ is changed to $\mathbf{B}'_d$ by modifying its PEs according to its allocated messages. Of course, messages can be decoded in the inverse order at the receiver's side. After $\mathbf{R}_d$, $\mathbf{G}_d$ and $\mathbf{B}_d$ finish embedding, we will modify white regions $\mathbf{R}_w$, $\mathbf{G}_w$ and $\mathbf{B}_w$ to embed messages the same way. Note that for each set, only N_h PEs with the highest smoothness are selected as host PEs, where N_h is usually six times the length of allocated messages, and the smoothness of a pixel is the variance of its four neighboring pixels (see Fig. 3).

Next, we compare the color RDH method under the proposed framework with two state-of-the-art color RDH schemes, the methods of Ou *et al.* [37] and Yao *et al.* [38]. As the proposed method is an extension of RCC from a consistent to inconsistent distortion metrics, we also compare the proposed method to RHM [20]. Note that the random scrambling PEs from the three color channels compose the single host sequence when performing RHM [20] on a color image. To show that the strength of the proposed method resulted from defining a multi-distortion metric, experiments on the proposed framework under a consistent distortion metric, i.e., $d_R(x,y) = d_G(x,y) = d_B(x,y) = (x-y)^2$, are also carried out for comparison.

As shown in Fig. 5, defining a multi-distortion metric for RCC is particularly meaningful, as it can greatly outperform the proposed framework under a consistent distortion metric and RHM [20] according to C-PSNR (Eq. (21). Furthermore, the proposed method outperforms those of Ou *et al.* [37] and

Yao *et al*. [38], implying that the proposed method is more suitable for color images due to inconsistent distortion metrics being reasonably defined by considering the characteristics of the human eye.

In addition to C-PSNR (Eq. (21)), as the grayscale version of a color image is rather valuable for many applications, to fully assess the performance of RDH in color image, we also convert both the host color image and the marked color image into grayscale images and calculate the PSNR between such two grayscale images, denoted by G-PSNR. Fig. 6 shows that the corresponding grayscale versions of marked images produced by the presented method are also of better visual quality than those produced by the existing methods [20], [37], [38].

Taking as examples the embedding of 50000 bits into typical $510 \times 510 \times 3$ color images, we also compare the speed of our algorithm to those of the existing color RDH methods; the embedding times (measured in seconds) are shown in Table I. Table I shows that our algorithm is even faster than the methods of Yao *et al*. and Ou *et al*. Note that in this paper, all test algorithms are implemented in MATLAB; speed comparison tests were run on a Lenovo personal computer with an i3-4130 CPU @ 3.40 GHz and 4.00 GB of RAM.

TABLE I: Speed comparisons of the proposed method and the existing color RDH algorithms.

Image	Lena	Baboon	Barbara	Airplane
Ou *et al*.'s method [37]	153 s	239 s	116 s	139 s
Yao *et al*.'s method [38]	97 s	96 s	95 s	94 s
Proposed method	91 s	91 s	86 s	85 s

B. Reversible steganography

Reversible steganography [31], [32] is a special kind of data hiding that has the reversibility of traditional RDH and the undetectability of traditional steganography. Such a technique is desired in covert storage [32] applications. To guarantee reversibility, reversible steganography cannot be as undetectable as traditional steganography. In this paper, we use the typical steganalyzer SPAM [40] to test undetectability, as done in [31] and [32]. All experiments are performed on the BOSSbase ver.1.01 [41] image database, which contains 10000 grayscale images of size 512×512. In the experiments, 5000 images are randomly selected for training, with the remaining 5000 images used for testing with ensemble classifiers [42]. We report the testing error, computed as the average of the false positive rate and false negative rate, randomly splitting the training and the testing images a total of 10 times. As for steganography, a higher detecting error rate implies a stronger undetectability.

Currently, the most successful steganographic approaches [25]–[28] are devoted to embedding messages while minimizing the total distortion, which is the sum of costs of all modified elements. In this paper, we define the cost for each element by HILL [27] as

$$\mathbf{C} = \frac{1}{|\mathbf{X} \bigotimes \mathbf{H}^{(1)}| \bigotimes \mathbf{L}_1} \bigotimes \mathbf{L}_2, \tag{27}$$

where $\mathbf{X} = (x_1, x_2, \cdots, x_N)$ is the input image, $\mathbf{C} = (c_1, c_2, \cdots, c_N)$ are the corresponding output cost values, $\mathbf{L}_1$ and $\mathbf{L}_2$ are two low-pass filters, and $\mathbf{H}^{(1)}$ is a high-pass filter (please refer to [27] for the details of HILL).

We also divide pixels of the host image into two parts, as shown in Fig. 3, and obtain the PE of a pixel by subtracting its prediction value, defined as the mean of its four neighboring pixels (see Fig. 3). If the PEs of dark pixels are used to carry messages, we will replace dark pixels with their corresponding prediction values to generate an interpolated image. Subsequently, we input such an interpolated image to calculate the cost value of each dark pixel using Eq. (27). White pixels will not be modified while performing RDH; thus, the recipient can recalculate the interpolated image and these cost values of dark pixels.

After obtaining the cost of each dark pixel, we select host PEs adaptively from dark regions according to the given payload. Specifically, for payload L, we select N_h PEs with the minimum cost values as host PEs, where

$$N_h = \min\{6L, \lfloor 0.4V \rfloor\}, \tag{28}$$

and V is the number of dark pixels. We can still enlarge N_h if the selected PEs are insufficient for accommodating the given payload. The theoretical motivation for this novel PE selection strategy is twofold: first, preferentially embedding messages into pixels with higher complexities, i.e., pixels with the smaller cost values, will increase the security of data hiding. Second, by scaling down the range of cost values, we can reduce the quantified loss when quantifying inconsistent distortion metrics as a multi-distortion metric.

As for the selected host PEs with the cost values $(c_1, c_2, \cdots, c_{N_h})$, we cluster the cost values into K classes by cluster algorithms such as K-means and denote the center of the ith class as C_i. Accordingly, the host PEs are divided into K subsequences, with the elements in the ith subsequence sharing the same distortion metric $d_i(x, y) = C_i|x - y|$, where $i = 1, 2, ..., K$. After the multi-distortion metric is defined, the presented framework is applied to minimize the total distortion.

As mentioned above, we divide the host sequence into K classes by cost values. It seems that with a larger K we can obtain a compound distortion metric that better fits the model of adaptive steganography. However, if the number of sub-histograms is large, the compound histogram will contain too many bins; hence, the entropy of the compound host sequence will decline with increasing K. As shown in Eq. (1), for RDH, the smaller the entropy of the host sequence is, the better the achievable performance of RDH. We can apply the following method to determine the suitable K for each host sequence. After the total distortion between the host sequence and the marked sequence denoted by J^K has been calculated for each class number K, the optimal K is determined by

$$K_{op} = arg \min_K J^K, K = 1, 2, 3, ... \quad . \tag{29}$$

A clear disadvantage of the above method is high computational complexity. The problem of obtaining the optimal number of classes not only adaptively but also effectively will be studied thoroughly in the future.

We provide examples using payloads of $L = 6000$ bits, $L = 8000$ bits and $L = 10000$ bits to show the influence of K on performance, using 1000 randomly selected images from BOSSbase ver.1.01 [41] as test images. It can be shown from Fig. 7, under the adopted HILL distortion metrics, that the total distortion by the presented framework reaches the minimum at $K = 3$. Therefore, $K = 3$ is adopted empirically for the steganalysis experiments in this subsection.

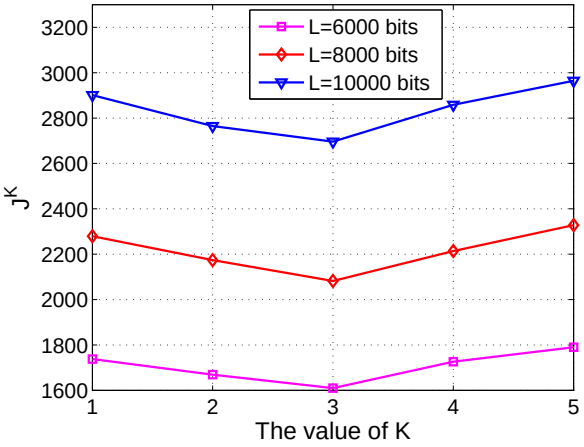


Fig. 7: The influence of K on the total distortion under different payloads.

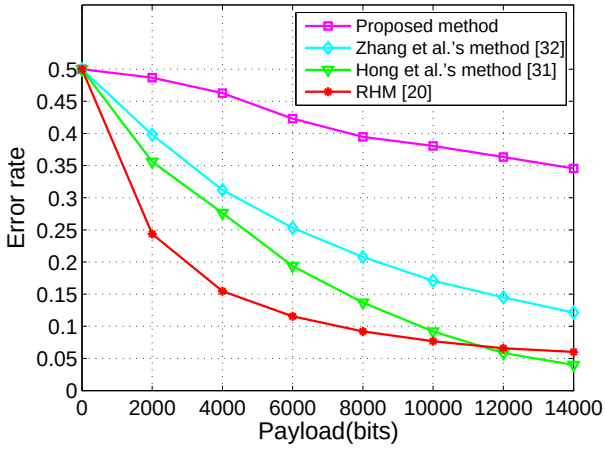


Fig. 8: Comparisons among the previous methods and the proposed method for resisting SPAM [40].

We compare the proposed method with two advanced reversible steganographic algorithms, the methods of Hong *et al.* [31] and Zhang *et al.* [32], under the detection by SPAM [40]. Zhang *et al.*'s RHM [20] is also added in the comparison experiments for contrast. As shown in Fig. 8, with the consistent distortion, it is difficult for RHM [20] to resist detection by SPAM. However, by extending RHM with the distortion metrics of adaptive steganography, the proposed method significantly outperforms the previous methods [20], [31], [32]. The power of our algorithm arises from defining the distortion metrics with the advanced HILL and minimizing the total distortion by the presented framework.

Taking as examples embedding 14000 bits into typical 510×510 grayscale images, the embedding times (measured in seconds) of our method and the previous reversible steganographic methods are listed in Table II. Table II shows that our algorithm is faster than that of Hong *et al.* [31] and slightly slower than those of Zhang *et al.* [32] and RHM [20]. Compared to the existing RCC schemes for a consistent distortion metric such as RHM [20], the additional operations by our method involve calculating the cost of each pixel and constructing the compound histogram and distortion metric, which usually have low computational complexity, as shown by Table II. Therefore, as an extended RCC framework, it does not introduce too much computational complexity in terms of the existing RCC schemes.

TABLE II: Comparison of the running times of our algorithm and the earlier reversible steganographic algorithms.

image	Lena	Baboon	Barbara	Airplane
Hong *et al.*'s method [31]	7 s	7 s	8 s	7 s
Zhang *et al.*'s method [32]	2 s	2 s	2 s	2 s
RHM [20]	1 s	1 s	1 s	4 s
Proposed method	2 s	3 s	2 s	7 s

V. Conclusions and Discussion

In this paper, we present the rate-distortion bound of RDH under a multi-distortion metric and develop a unified framework for estimating the optimal transition probability matrix under a multi-distortion metric that enables us to extend recursive code construction schemes to applications of inconsistent distortion metrics. The experiments demonstrate that the proposed method significantly outperforms the previous methods.

Inconsistent distortion metrics are quantified into a multi-distortion metric when performing reversible steganography. As discussed above, increasing the number of sub-histograms is not necessarily beneficial. Obtaining the optimal number of classes not only adaptively but also effectively is a difficult theoretical problem to be solved in the future. On the other hand, we apply HILL to define the cost of modification for each pixel; however, it is clear that such distortion metrics used in steganography should not be applied directly to reversible steganography. In the future, we will design special distortion metrics for reversible steganography and then design more secure RDH algorithms.

VI. Acknowledgments

The authors thank Ou *et al.* and Yao *et al.* for offering the source codes in their papers [37], [38]. To help readers apply the proposed framework, we will post the MATLAB implementation of this paper on our website at http://home.ustc.edu.cn/%7Ehoudd.

Appendix

We prove that the optimal solution to Problem II can provide the optimal solution to Problem I both theoretically and experimentally as follows.

Assume $\mathbf{P}_{Y_c|X_c}$ in the form of Eq. (17) to be the optimal solution of Problem II, and $\mathbf{P}_{Y_c}$ given by Eq. (13) is the corresponding optimal distribution of the marked sequence. This solution reaches capacity R such that $H(Y_C) - H(X_C) = R$. We denote the average distortion achieved by this solution as

$$J_{comp} = \sum_{x=0}^{Km-1} \sum_{y=0}^{Kn-1} P_{X_c}(x) P_{Y_c|X_c}(y|x) d_c(x,y). \quad (30)$$

With the optimal solution $\mathbf{P}_{Y_c|X_c}$, we can construct a solution of Problem I such that

$$\mathbf{P}_{Y_i|X_i} = \mathbf{P}_{Y_i^t|X_i^t}, 1 \le i \le K. \quad (31)$$

From Eq. (12), $\mathbf{P}_{X_i} = \frac{N}{N_i}\mathbf{P}_{X_i^t}$; hence, Eq. (31) implies that the corresponding optimal marginal distribution of the marked sequence satisfies

$$\mathbf{P}_{Y_i} = \frac{N}{N_i}\mathbf{P}_{Y_i^t}. \quad (32)$$

The solution Eq. (31) can reach capacity R because

$$\begin{aligned} &\frac{\sum_{i=1}^K N_i \times H(Y_i)}{N} - \frac{\sum_{i=1}^K N_i \times H(X_i)}{N} \\ &= H(Y_C) - H(X_C) \\ &= R \ . \end{aligned} \quad (33)$$

Additionally, the corresponding average distortion J_{mult} satisfies

$$\begin{aligned} J_{mult} &= \frac{\sum_{i=1}^K N_i \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} P_{X_i}(x) P_{Y_i|X_i}(y|x) d_i(x,y)}{N} \\ &= \sum_{x=0}^{Km-1} \sum_{y=0}^{Kn-1} P_{X_c}(x) P_{Y_c|X_c}(y|x) d_c(x,y) \\ &= J_{comp} \ . \end{aligned} \quad (34)$$

Next, we prove that the solution in Eq. (31) is the optimal solution of Problem I. If it is not, there exists another solution of Problem I that can reach capacity R and achieve a smaller average distortion. We denote such a solution as

$$\mathbf{P}^*_{Y_i|X_i}, 1 \le i \le K \quad (35)$$

and the corresponding marginal distribution of the marked sequence as $\mathbf{P}^*_{Y_i}$ for $1 \le i \le K$. The average distortion achieved by the solution in Eq. (35) is

$$J^*_{mult} = \frac{\sum_{i=1}^K N_i \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} P_{X_i}(x) P^*_{Y_i|X_i}(y|x) d_i(x,y)}{N}, \quad (36)$$

satisfying $J^*_{mult} < J_{mult}$.

Using the solution in Eq. (35), we can construct a solution for Problem II. Define X_i^t and Y_i^t by adding offsets ox_i and oy_i to the ranges of X_i and Y_i, respectively, and define

$$\mathbf{P}^*_{Y_i^t|X_i^t} = \mathbf{P}^*_{Y_i|X_i}, 1 \le i \le K. \quad (37)$$

Furthermore, define

$$\mathbf{P}^*_{Y_c|X_c} = \begin{bmatrix} \mathbf{P}^*_{Y_1^t|X_1^t} & 0 & \dots & 0 \\ 0 & \mathbf{P}^*_{Y_2^t|X_2^t} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \mathbf{P}^*_{Y_K^t|X_K^t} \end{bmatrix}. \quad (38)$$

With Eq. (12) and Eq. (37), the corresponding marginal distribution is given by $\mathbf{P}^*_{Y_i^t} = \frac{N_i}{N}\mathbf{P}^*_{Y_i}$, and

$$\mathbf{P}^*_{Y_C} = [\mathbf{P}^*_{Y_1^t}, \cdots, \mathbf{P}^*_{Y_K^t}]. \quad (39)$$

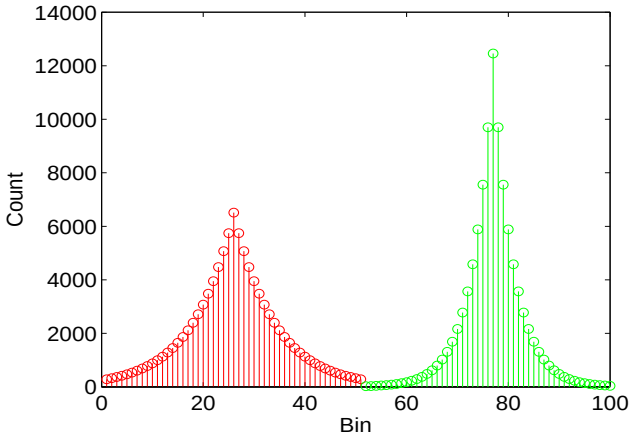


Fig. 9: Two generated subsequences following the Laplace distribution.

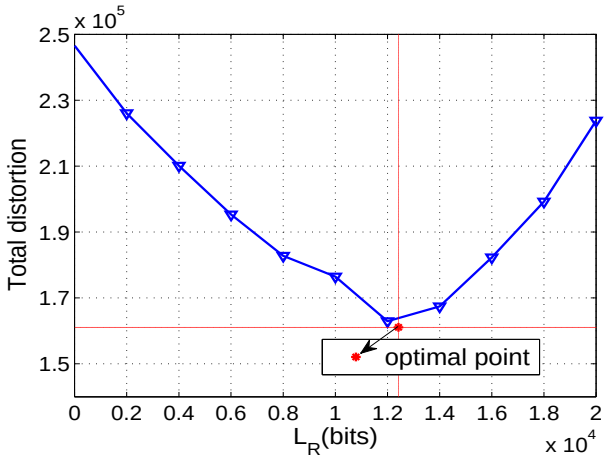


Fig. 10: Searching for the optimal strategy of payload allocation causing the minimum distortion.

Clearly, the solution in Eq. (38) to Problem II can reach the same capacity R because

$$\begin{aligned} &H(Y_C^*) - H(X_C) \\ &= \frac{\sum_{i=1}^K N_i \times H(Y_i^*)}{N} - \frac{\sum_{i=1}^K N_i \times H(X_i)}{N} \\ &= R \ . \end{aligned} \quad (40)$$

However, the average distortion J^*_{comp} achieved by the solution in Eq. (38) satisfies

$$\begin{aligned} J^*_{comp} &= \sum_{x=0}^{Km-1} \sum_{y=0}^{Kn-1} P_{X_c}(x) P^*_{Y_c|X_c}(y|x) d_c(x,y) \\ &= \frac{\sum_{i=1}^K N_i \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} P_{X_i}(x) P^*_{Y_i|X_i}(y|x) d_i(x,y)}{N} \\ &= J^*_{mult} < J_{mult} = J_{comp} \ . \end{aligned} \quad (41)$$

This is contrary to the assumption that $\mathbf{P}_{Y_C|X_C}$ is the optimal solution of Problem II. Thus, we have proven that the solution Eq. (31) is the optimal solution of Problem I.

In fact, following a similar method, we can also prove that the optimal solution of Problem I can yield the optimal solution of Problem II.

The optimality of the proposed method can be illustrated with a simple example. Two subsequences following different Laplace distributions are generated with the range [0, 49] and the same length of 100000, and the corresponding histograms are depicted in red and green in Fig. 9. The corresponding distortion metrics are defined as $d_R = (x-y)^2$ and $d_G = 2^2(x-y)^2$, respectively. Assume that the length of the message to be embedded is $L = 20000$ bits. We can estimate the optimal strategy of payload allocation between the two subsequences by an exhaustive search. Assume that the payload for the red subsequence is L_R bits; hence, the remaining payload for the green subsequence is $L - L_R$ bits. We try a series of L_R with the step length of 2000 bits to estimate the optimal solution that can minimize the total distortion. As shown in Fig. 10, the total distortion reaches the minimal value at L_R close to 12000 bits. Clearly, the computational complexity of finding the optimal solution by an exhaustive search will be too high if the number of subsequences increases. However, using the proposed method, the optimal point, represented by the red point in Fig. 10, can be solved for easily.

REFERENCES

[1] A. S. Brar, M. Kaur, "Reversible Watermarking Techniques for Medical Images with ROI-Temper Detection and Recovery - A Survey," International Journal of Emerging Technology and Advanced Engineering, vol. 2, no. 1, pp. 32-36, Jan. 2012.

[2] J. Fridrich and M. Goljan, "Lossless Data Embedding for All Image Formats," in SPIE Proceedings of Photonics West, Electronic Imaging, Security and Watermarking of Multimedia Contents, vol. 4675, pp. 572-583, San Jose, Jan. 2002.

[3] K. Chung, Y. Huang, P. Chang, *et al.*, "Reversible Data Hiding-Based Approach for Intra-Frame Error Concealment in H.264/AVC," IEEE Trans. Circuits System and Video Technology, vol. 20, no. 11, pp. 1643-1647, Nov. 2010.

[4] D. Hou, W. Zhang, Z. Zhan, *et al.*, "Reversible image processing via reversible data hiding", in 2016 IEEE International Conference on Digital Signal Processing (DSP), pp. 427-431, 2016.

[5] J. Tian, "Reversible Data Embedding Using a Difference Expansion," IEEE Trans. Circuits System and Video Technology, vol. 13, no. 8, pp. 890-896, Aug. 2003.

[6] D. Thodi and J. Rodriguez, "Expansion Embedding Techniques for Reversible Watermarking," IEEE Trans. Image Processing, vol. 16, no. 3, pp. 721-730, Mar. 2007.

[7] Y. Hu, H. Lee, and J. Li, "DE-based Reversible Data Hiding with Improved Overflow Location Map," IEEE Trans. Circuits System and Video Technology, vol. 19, no. 2, pp. 250-260, Feb. 2009.

[8] Z. Ni, Y. Shi, N. Ansari, and S. Wei, "Reversible Data Hiding," IEEE Trans. Circuits System and Video Technology, vol. 16, no. 3, pp. 354-362, 2006.

[9] P. Tsai, Y.C. Hu, and H.L. Yeh, "Reversible Image Hiding Scheme Using Predictive Coding and Histogram Shifting," Signal Processing, vol.89, pp. 1129-1143, 2009.

[10] V. Sachnev, H. J. Kim, J. Nam, S. Suresh, and Y. Shi, "Reversible Watermarking Algorithm Using Sorting and Prediction," IEEE Trans. Circuits and Systems for Video Technology, vol. 19, no. 7, pp. 989-999, July 2009.

[11] L. Luo, Z. Chen, M. Chen, *et al.*, "Reversible Image Watermarking Using Interpolation Technique," IEEE Trans. Information Forensics and Security, vol. 5, no. 1, pp. 187-193, Mar. 2010.

[12] F. Peng, X. Li and B. Yang, "Adaptive Reversible Data Hiding Scheme Based on Integer Transform," Signal Processing, vol. 92, no. 1, pp. 54-62, Jan. 2012.

[13] X. Li, B. Yang, and T. Zeng, "Efficient Reversible Watermarking Based on Adaptive Prediction-Error Expansion and Pixel Selection," IEEE Trans. Image Processing, vol. 20, no. 12, pp. 3524-3533, Dec. 2011.

[14] B. Ou, X. Li, J. Wang and F. Peng, "High-fidelity reversible data hiding based on geodesic path and pairwise prediction-error expansion," Neurocomputing, vol. 226, pp. 23-34 , Feb, 2017.

[15] J. Wang, J. Ni, X. Zhang and Y. Shi, "Rate and distortion optimization for reversible data hiding using multiple histogram shifting," IEEE Trans. cybernetics, vol.47, no.2, pp. 315-326, Feb, 2017.

[16] T. Kalker, F. M. Willems, "Capacity Bounds and Code Constructions for Reversible Data-Hiding," in Proc. of 14th International Conference on Digital Signal Processing (DSP2002), pp. 71-76, 2002.

[17] S-J. Lin, and W-H Chung, "The Scalar Scheme for Reversible Information-Embedding in Gray-scale Signals:Capacity Evaluation and Code Constructions," IEEE Trans. Information Forensics and Security, vol. 7, no. 4, pp. 1155-1167, April 2012.

[18] W. Zhang, B. Chen, and N. Yu, "Capacity-Approaching Codes for Reversible Data Hiding," in Proc. of 13th Information Hiding Conference, Prague, May 2011.

[19] W. Zhang, B. Chen, and N. Yu, "Improving Various Reversible Data Hiding Schemes via Optimal Codes for Binary Covers," IEEE Trans. Image Processing, Vol. 21, no. 6, pp. 2991-3003. 2012.

[20] 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 Processing, Vol. 22, no. 7, pp. 2775-2785, July 2013.

[21] X. Zhang, "Reversible data hiding with optimal value transfer," IEEE Trans. Multimedia, Vol. 15, no. 2, pp. 316-325, Feb. 2013.

[22] F. Willems, D. Maas, and T. Kalker, "Semantic lossless source coding," in Proc. 42nd Ann. Allerton Conf. Commun., Control and Comput., Monticello, IL, USA, 2004.

[23] X. Hu, W. Zhang, X. Hu, N. Yu, X. Zhao, and F. Li, "Fast Estimation of Optimal Marked-Signal Distribution for Reversible Data Hiding," IEEE Trans. Information Forensics and Security, vol. 8, no. 5, pp. 779-788, May 2013.

[24] W. Zhang, X. Hu, X. Li, and Y. Nenghai, "Optimal transition probability of reversible data hiding for general distortion metrics and its applications," IEEE Trans. Image Processing, vol. 24, no. 1, pp. 294-304, 2015.

[25] V. Holub and J. Fridrich, "Designing steganographic distortion using directional filters," in *2012 IEEE International Workshop on Information Forensics and Security (WIFS)*, pp. 234-239, Dec 2012.

[26] V. Holub, J. Fridrich, and T. Denemark, "Universal distortion function for steganography in an arbitrary domain," *EURASIP Journal on Information Security*, vol. 2014, no. 1, pp. 1-13, 2014.

[27] B. Li, M. Wang, J. Huang, and X. Li, "A new cost function for spatial image steganography," in *2014 IEEE International Conference on Image Processing (ICIP)*, pp. 4206-4210, 2014.

[28] B. Li, M. Wang, X. Li, S. Tan, and J. Huang, "A strategy of clustering modification directions in spatial image steganography," *IEEE Trans. Information Forensics and Security*, vol. 10, no. 9, pp. 1905-1917, 2015.

[29] S.-W. Jung, S.-J. Ko, *et al.*, "A new histogram modification based reversible data hiding algorithm considering the human visual system," *IEEE Signal Processing Letters*, vol. 18, no. 2, pp. 95-98, 2011.

[30] Y. Yang, W. Zhang, X. Hu, and N. Yu, "Improving visual quality of reversible data hiding by twice sorting," *Multimedia Tools and Applications*, vol. 75, no. 21, pp. 13663-13678, 2016.

[31] W. Hong, T. Chen, J. Chen. "Reversible data hiding using Delaunay triangulation and selective embedment," Information Sciences, vol. 308, pp. 140-154, 2015.

[32] Z. Zhang and W. Zhang, "Reversible steganography: Data hiding for covert storage," in *2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, pp. 753-756, 2015.

[33] Rec. 601: https://en.wikipedia.org/wiki/Rec.%20601.

[34] X. Hu, W. Zhang, X. Li, and N. Yu, "Minimum rate prediction and optimized histograms modification for reversible data hiding," IEEE Trans. Information Forensics and Security. vol. 10, no. 3, pp. 653-664, 2015.

[35] X. Li, W. Zhang, X. Gui, and B. Yang, "Efficient reversible data hiding based on multiple histograms modification," *IEEE Trans. Information Forensics and Security*, vol. 10, no. 9, pp. 2016-2027, 2015.

[36] J. Li, X. Li, and B. Yang, "Reversible data hiding scheme for color image based on prediction-error expansion and cross-channel correlation," Signal Processing. vol. 93, no. 9, pp. 2748-2758, 2013.

[37] B. Ou, X. Li, Y. Zhao, and R. Ni, “Efficient color image reversible data hiding based on channel-dependent payload partition and adaptive embedding,” *Signal Processing*. vol. 108, pp. 642-657, 2015.
[38] H. Yao, C. Qin, Z. Tang, and Y. Tian, “Guided filtering based color image reversible data hiding,” Journal of Visual Communication and Image Representation, vol. 43, pp. 152-163, 2017.
[39] K. He, J. Sun, X. Tang, “Guided image filtering,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 35, no. 6, pp. 1397 - 1409 , June 2013.
[40] T. Pevny, P. Bas, and J. Fridrich, “Steganalysis by subtractive pixel adjacency matrix,” *IEEE Trans. Information Forensics and Security*, vol. 5, no. 2, pp. 215-224, 2010.
[41] T Filler, T Pevny, P Bas, BOSS (Break Our Steganography System). http://www.agents.cz/boss, accessed date 20/12/13.
[42] J. Kodovsky, J. Fridrich, V. Holub, “Ensemble classifiers for steganalysis of digital media,” IEEE Trans. Information Forensics and Security, vol. 7, no. 2, pp. 432-444, 2012.