Optimal Transition Probability of Reversible Data Hiding for General Distortion Metrics and Its Applications

Weiming Zhang, Xiaocheng Hu, Xiaolong Li, and Yu Nenghai

***Abstract*— Recently, a recursive code construction (RCC) approaching the rate-distortion bound of reversible data hiding (RDH) was proposed. However, to estimate the rate-distortion bound or execute RCC, one should first estimate the optimal transition probability matrix (OTPM). By previous methods, OTPM can be effectively estimated only for some specific distortion metrics, such as square error distortion or L_1-Norm. In this paper, we proposed a unified framework of estimating the OTPM for general distortion metrics, with which we can calculate the rate-distortion bound of RDH for general cases and extend RCC to improve state-of-the-art RDH schemes based on any distortion metrics.**

***Index Terms*— Reversible data hiding, embedding rate, distortion, recursive code construction, Lagrange duality, convex optimization.**

I. Introduction

REVERSIBLE data hiding (RDH) [1] is a technique embedding data into a cover, by which the original cover can be losslessly restored after the embedded data is extracted. This important technique is widely used in medical imagery [2], military imagery [3] and law forensics, where no distortion of the original cover is allowed. It has been found that RDH can also be helpful in video error-concealment coding [4].

Until now, many RDH techniques have been proposed based on three fundamental strategies: lossless compression-appending scheme [3], difference expansion (DE) [5]–[7] and histogram shift (HS) [8]. Some recent arts combined these three strategies to residuals of the image such as prediction errors (PE) [9]–[13] to achieve better performance.

Almost all state-of-the-art RDH algorithms consist of two steps. The first step generates a host sequence with small entropy, i.e., the host has a sharp histogram which usually can be realized by using PE combined with the sorting technique [10] or pixel selection [13]. The second step reversibly embeds the message into the host sequence by modifying its histogram with methods like HS and DE.

One natural problem is what is the upper bound of the payload for a given host sequence and a distortion constraint. For i.i.d. host sequence, this problem has been solved by Kalker and Willems [14], who formulated the RDH as a special rate-distortion problem, and obtained the rate-distortion function, i.e., the upper bound of 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 host signal and marked signal respectively. The maximum entropy is over all transition probability matrices $P_{Y|X}(y|x)$ satisfying the distortion constraint $\sum_{x,y} P_X(x)P_{Y|X}(y|x)D(x,y) \le \Delta$. The distortion metric $D(x, y)$ is usually defined as the square error distortion, $D(x, y) = (x - y)^2$.

Therefore, to evaluate the capacity of RDH, one should calculate the optimal transition probability matrix (OTPM) $P_{Y|X}(y|x)$. For some specific distortion metrics $D(x, y)$, such as square error distortion $D_s(x, y) = (x - y)^2$ or L_1-Norm $D_1(x, y) = |x - y|$, the OTPM has a Non-Crossing-Edges (NCE) property [15]. Using this NCE property, the optimal solution on $P_{Y|X}(y|x)$ can be analytically derived by the marginal distributions $P_X(x)$ and $P_Y(y)$. Because $P_X(x)$ is given, for distortion metrics satisfying NCE property, the problem is how to calculate the optimal $P_Y(y)$. Lin et al. [15] proposed a method to estimate the optimal marginal distribution $P_Y(y)$, by which they can evaluate the capacity (1) for distortion metrics, such as square error distortion or L_1-Norm. In [16], we proposed a fast algorithm to estimate the optimal marginal distribution $P_Y(y)$ for both the distortion constrained problem (1) and its dual problem, i.e., the embedding rate constrained problem.

On the other hand, the OTPM implies the optimal modification manner on the histogram of the host signal X. According to the OTPM, Lin et al. [15] proposed a coding method that can be close to the rate-distortion bound (1). By improving the recursive code construction (RCC) [14], we obtain the optimal embedding methods of RDH for binary host sequences [17], [18] and general gray-scale host sequences [19] respectively, and meanwhile we proved that RCC can approach the rate-distortion bound (1) as long as the entropy coder reaches entropy, which establishes the

Manuscript received January 23, 2014; revised May 21, 2014 and September 3, 2014; accepted September 8, 2014. Date of publication September 17, 2014; date of current version December 16, 2014. This work was supported in part by the Natural Science Foundation of China under Grant 61170234 and Grant 60803155, in part by the Strategic Priority Research Program, Chinese Academy of Sciences, Beijing, China, under Grant XDA06030601, and in part by the Science and Technology on Information Assurance Laboratory under Grant KJ-13-02. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Hitoshi Kiya.

W. Zhang, X. Hu, and Y. Nenghai are with the CAS Key Laboratory of Electromagnetic Space Information, University of Science and Technology of China, Hefei 230026, China (e-mail: zhangwm@ustc.edu.cn; hxc@mail.ustc.edu.cn; ynh@ustc.edu.cn).

X. Li is with the Institute of Computer Science and Technology, Peking University, Beijing 100871, China (e-mail: lixiaolong@icst.pku.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2014.2358881

equivalency between RDH and lossless data compression. A coding method similar to [19] is independently presented by Zhang [20].

The coding and decoding processes of the methods in [15], [19], and [20] need the OTPM as parameters. In other words, to evaluate the capacity or to constructe capacity-approaching codes for RDH, one should first solve the optimization problem (1). As mentioned above, for distortion metrics satisfying the NCE property, this problem has been solved in [15] and [16]. But for other distortion metrics, such as Hamming distance that is useful for RDH in binary images, the NCE property no longer holds and the OTPM can not be obtained analytically.

Recently, Ou et al. [21] proposed an efficient RDH method by estimating pairwise prediction-errors and modifying the 2D histogram of the PE, which outperforms previous RDH methods. However, in the 2D histogram case, the distortion metric is Euclidean distance that does not satisfy the NCE property, so we cannot estimate its capacity with methods in [15] and [16], nor apply the optimal code constructions to it. Another example of RDH, that does not meet the condition of NCE, is the Pattern Substitution method [22] for binary images, in which the distortion metric is a Hamming-like distance.

In summary, although RCC can minimize the embedding distortion of RDH, one should first estimate the OTPM. Previous methods can only estimated the OTPM for some specific distortion metrics, such as square error distortion and L1-Norm, which limits the applications of RCC. In the present paper, we proposed a unified framework for estimating OTPM for general distortion metrics, with which we can calculate the rate-distortion bound of RDH for general cases and can extend the RCC method [19] to various scenarios. Taking the RDH schemes in [21] and [22] as examples, we show how to apply the extended RCC to improve existing RDH schemes with any distortion metrics.

The rest of the paper is organized as follows. In Section II, we briefly introduce how to estimate OPTM with NCE property and how to embed data with RCC. The method on estimating OTPM for generic distortion metrics is elaborated in Section III, and a improvement method on RCC is presented in Section IV. In Section V, two application cases are demonstrated to show the power of the proposed method under different kinds of distortion metrics. The paper is concluded with a discussion in Section VI.

II. Previous Arts

Throughout this paper, we denote matrices and vectors by boldface fonts, and use capital letters for the random variables and small letters for their realizations. We denote the entropy by $H(X)$ and the conditional entropy by $H(Y|X)$. Specially, for the probability distribution $(P(0), \ldots, P(B-1))$ such that $P(i) > 0$ and $\Sigma_{i=0}^{B-1} P(i) = 1$, the B-ary entropy function is defined as $H(P(0), \ldots, P(B-1)) = -\Sigma_{i=0}^{B-1} P(i) \log_2 P(i)$.

A. Optimal Transition Probability With NCE Property

In RDH, L bits of message $\mathbf{m} = (m_1, \ldots, m_L)$ are embedded by the sender into the cover sequence $\mathbf{x} = (x_1, \ldots, x_n) \in \mathcal{X}^n$, through slightly modifying its elements to produce a marked-sequence $\mathbf{y} = (y_1, \ldots, y_n) \in \mathcal{Y}^n$. The two alphabet sets, $\mathcal{X} = \{0, 1, \ldots, M-1\}$ and $\mathcal{Y} = \{0, 1, \ldots, N-1\}$, are both finite. We denote the embedding rate by $R = L/n$. Schemes are usually constructed to minimize some distortion metric $D(x, y)$ between $\mathbf{x}$ and $\mathbf{y}$ for a given embedding rate R. The distortion metric $D(x, y)$ could be the square error distortion $D_s(x, y) = (x-y)^2$, the L_1-Norm distortion $D_1(x, y) = |x - y|$, or a user defined distortion, etc. We assume the cover $\mathbf{x} = (x_1, \ldots, x_n)$ is an n-tuple composed of n i.i.d. samples drawn from the probability distribution $P_X = \{P_X(x), x \in \mathcal{X}\}$.

According to Eq. (1), given an embedding rate R, we get a corresponding constraint condition on $H(Y)$ that is denoted as H_Y. In other words, $R = H_Y - H(X)$. The problem to minimize the average distortion for a given embedding rate $R = H_Y - H(X)$ can be formulated as:

$$\begin{aligned} \text{minimize} \quad & \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} P_X(x) P_{Y|X}(y|x) D(x, y) \\ \text{subject to} \quad & -\sum_{y=0}^{N-1} P_Y(y) \log_2(P_Y(y)) \geq H_Y \\ & \sum_{x=0}^{M-1} P_X(x) P_{Y|X}(y|x) = P_Y(y), \forall y \\ & \sum_{y=0}^{N-1} P_{Y|X}(y|x) = 1, \forall x \\ & P_{Y|X}(y|x) \geq 0, \forall x, y \end{aligned} \tag{2}$$

where the variables are the transition probability matrix $P_{Y|X}(y|x)$, the constant parameters are the cover distribution $P_X(x)$, the distortion measure matrix $D(x, y)$, and the entropy constraint H_Y.

In [16], we proved that the Lagrange dual [23] of the above rate distortion problem (2) is of the following form:

$$\begin{aligned} \text{minimize} \quad & \gamma \sum_{j=0}^{N-1} e^{v_j/\gamma - 1} + \sum_{i=0}^{M-1} u_i - \gamma H_Y \\ \text{subject to} \quad & P_X(i) D(i, j) + u_i + P_X(i) v_j \geq 0, \quad \forall i, j \gamma \geq 0 \end{aligned} \tag{3}$$

where the dual variables are $\mathbf{v} \in \mathbb{R}^N$, $\mathbf{u} \in \mathbb{R}^M$, and $\gamma \in \mathbb{R}$. The constant parameters are $P_X(i)$, $D(i, j)$, and H_Y.

The optimal solutions v_j and γ of the dual problem (3) yield the optimal marked-signal distribution of the primal problem (2) with the following form:

$$P_Y(j) = e^{v_j/\gamma - 1}, j = 0, \ldots, N-1. \tag{4}$$

The optimal objective function value of (2), i.e., the average distortion D_{av}, is given by:

$$D_{av} = -(\gamma \sum_{j=0}^{N-1} e^{v_j/\gamma - 1} + \sum_{i=0}^{M-1} u_i - \gamma H_Y). \tag{5}$$

For some specific distortion metrics $D(x, y)$, like square error distortion $D_s(x, y) = (x-y)^2$ or L_1-Norm distortion

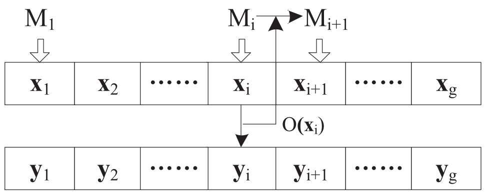


Fig. 1. Illustration of recursive code construction.

$D_1(x, y) = |x - y|$, it has been proved both in [15] and [24] that the OTPM $P_{Y|X}(y|x)$ has the Non-Crossing-Edges (NCE) property. To be specific, the NCE property means that, given an optimal $P^*_{Y|X}$, for any two distinct possible transition events $P^*_{Y|X}(y_1|x_1) > 0$ and $P^*_{Y|X}(y_2|x_2) > 0$, if $x_1 < x_2$, then $y_1 \leq y_2$ holds. Lin et al. [15] pointed out that, by using the NCE property, the joint distribution of X and Y can be expressed by (6),

$$P_{X,Y}(x, y) = \max\{0, \min\{P_{CX}(x), P_{CY}(y)\} - \max\{P_{CX}(x-1), P_{CY}(y-1)\}\}, \quad (6)$$

where $P_{CX}(x)$ and $P_{CY}(y)$ are cumulative probability distributions of X and Y defined by $P_{CX}(x) = \sum_{i=0}^{x} P_X(i)$, $x = 0, \ldots, M-1$, and $P_{CY}(y) = \sum_{i=0}^{y} P_Y(i)$, $y = 0, \ldots, N-1$. Noted that $P_{CX}(-1) = P_{CY}(-1) = 0$ and $P_{CX}(M-1) = P_{CY}(N-1) = 1$.

Therefore, when satisfying NCE property, the OTPM can also be deduced analytically by the marginal distributions $P_X(x)$ and $P_Y(y)$. However, for other distortion metrics, such as Hamming distortion, or even user defined distortions, the NCE property may no longer hold and then OTPM cannot be obtained analytically.

B. Recursive Code Construction

After obtaining optimal $P_{Y|X}(y|x)$, we can calculate the other OTPM $P_{X|Y}(x|y)$ because P_X is given. With optimal $P_{Y|X}(y|x)$ and $P_{X|Y}(x|y)$, we proposed a capacity-approaching coding method called recursive code construction (RCC). By RCC, we divide the host sequence into disjoint blocks and embed the message by modifying the histogram of each block.

As the RCC behaves exactly the same within each block, we take a single block to illustrate the data embedding process. As shown in Fig. 1, assume the block sequence $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{iK})$ is a K-tuple composed of K samples drawn with probability distribution P_X. According to the probability transition matrix $P_{Y|X}(y|x)$, we on average can decompress $S = KH(Y|X)$ bits of message $\mathbf{M}_i = (m_{i1}, m_{i2}, \ldots, m_{iS})$ into a marked-sequence $\mathbf{y}_i = (y_{i1}, y_{i2}, \ldots, y_{iK})$. To restoring $\mathbf{x}_i$ from $\mathbf{y}_i$, we compress $\mathbf{x}_i$ according to $P_{X|Y}(x|y)$, and the compressed information $O(\mathbf{x}_i)$ in consort with $\mathbf{M}_{i+1}$ are embedded into the next block. The average length of $O(\mathbf{x}_i)$ is $KH(X|Y)$, so the pure average payload of each cover signal is equal to

$$\frac{1}{K}[KH(Y|X) - KH(X|Y)] = H(Y|X) - H(X|Y) = H(Y) - H(X). \quad (7)$$

Therefore, when using OTPM, RCC can approach the rate-distortion bound (1).

The data extraction and cover restoration are executed in a backward manner. After extracting $O(\mathbf{x}_i)$, the cover block $\mathbf{x}_i$ can be restored by decompressing $O(\mathbf{x}_i)$ according to $P_{X|Y}(x|y)$ with the help of $\mathbf{y}_i$, and then $\mathbf{M}_i$ can be extracted by compressing $\mathbf{y}_i$ according to $P_{Y|X}(y|x)$ with the help of $\mathbf{x}_i$. For details of the RCC method, we refer readers to [19].

RCC can minimize the embedding distortion of RDH. With RCC, designers of RDH only need to consider how to generate a host sequence having small entropy and how to estimate the OTPM for a given embedding rate.

III. Optimal Transition Probability With General Distortion Metrics

As mentioned before, previous method can only estimate OTPM for distortion metrics satisfying NCE property. In this section, we focus on calculating the OTPM from the estimated optimal dual variables of problem (3) for general distortion metrics.

In [16], the hidden Lagrange multiplier $\Lambda = \{\lambda_{x,y}|\forall x, y\} \in \mathbb{R}^{M \times N}$ for the dual problem (3) associated with $\mathbf{u}$ and $\mathbf{v}$ is given by

$$\lambda_{x,y} = P_X(x)D(x, y) + u_x + P_X(x)v_y, \forall x, y \quad (8)$$

From the complementary slackness [23], the primal OTPM $P_{Y|X}(y|x)$ and the dual optimal variable Λ satisfy the condition

$$\lambda_{x,y} P_{Y|X}(y|x) = 0, \; x = 1, 2, \ldots, M, \; y = 1, 2, \ldots, N, \quad (9)$$

which means that the OTPM $P_{Y|X}(y|x)$ is a sparse matrix whose nonzero elements only occur where $\lambda_{x,y} = 0$. As a result, the optimal $P_{Y|X}(y|x)$ satisfies the following non-negative linear equations:

$$\begin{aligned} \sum_{x,y} P_X(x) P_{Y|X}(y|x) D(x, y) &= D_{av} \\ \sum_{x=0}^{M-1} P_X(x) P_{Y|X}(y|x) &= P_Y(y), \forall y \\ \sum_{y=0}^{N-1} P_{Y|X}(y|x) &= 1, \forall x \\ P_{Y|X}(y|x) &\geq 0, \forall x, y \\ \lambda_{x,y} P_{Y|X}(y|x) &= 0, \forall x, y \end{aligned} \quad (10)$$

As the complementary slackness (9) shows that $P_{Y|X}(y|x) = 0$ if $\lambda_{x,y} > 0$, we immediately obtain the support of the optimal $P_{Y|X}(y|x)$ as

$$I_{xy} = \{(x, y)|\lambda_{x,y} = 0\}. \quad (11)$$

From I_{xy}, all the unknown elements in $P_{Y|X}(y|x)$ are gathered by $\{P_{Y|X}(y|x), (x, y) \in I_{xy}\}$. We arrange these unknown variables as a column vector $\mathbf{p} = (p_1, p_2, \ldots, p_s)$ where s is the cardinality of support set I_{xy}.

By reformulating the linear equations in (10), the column vector $\mathbf{p}$ satisfies a reduced system of linear equations:

$$\begin{aligned} A\mathbf{p} &= b \\ \mathbf{p} &\geq 0, \end{aligned} \tag{12}$$

where $A \in \mathbb{R}^{r \times s}$, $r = M + N + 1$ is the number of linear equations, and s is the number of unknowns. The nature of the solution depends upon the rank of the matrix A.

- CASE 1: $rank(A) = s$
  In this case, r equations in A are linearly independent, which yields a single solution to the following equations:
  $$A\mathbf{p} = b. \tag{13}$$
  The solution of $\mathbf{p}$ is straightforward by solving (13), regardless of the non-negative constraint in (12).
- CASE 2: $rank(A) < s$
  In this case, the number of linearly independent equations in A is less than the number of unknowns, which means the system of linear equations (13) is underdetermined. So among the solutions of (13), we have to chase the one that follows the non-negative constraint in (12). This can be solved by the following linear program:
  $$\begin{aligned} \text{maximize} \quad & \sum_{i=1}^{s} p_i \\ \text{subject to} \quad & A\mathbf{p} = b \\ & \mathbf{p} \geq 0. \end{aligned} \tag{14}$$

The solution set of $\mathbf{p}$ depends on the distortion metric $D(x, y)$. For some distortion metrics, such as the square error distortion $D_s(x, y) = (x-y)^2$ or L_1-Norm $D_1(x, y) = |x-y|$, a single analytic solution exists, while for other distortion metrics, there may exist multiple solutions in (14).

So, for the general distortion metric $D(x, y)$, we use a two-step strategy to estimate the optimal $P_{Y|X}(y|x)$. Firstly, we use the fast algorithm in [16] to get the marginal distribution $P_Y(y)$ and the corresponding dual variables. Secondly, we solve the above linear equations (13) or the linear program (14) to finally estimate the optimal transition probability matrix $P_{Y|X}(y|x)$.

In the following example, we show the optimal transition probability matrix obtained by the above two-step strategy for the square error distortion.

Example 1: Consider the square error distortion case, where $D_s(x, y) = (x - y)^2$. We take a simple distribution $P_X = (0.1, 0.3, 0.2, 0.1, 0.1, 0.1, 0.05, 0.05)$, the two alphabets are $\mathcal{X} = \mathcal{Y} = \{0, 1, 2, 3, 4, 5, 6, 7\}$. To embed $H_Y - H(X) = 0.2$ bpp (bits per pixel) message, using the above two-step strategy, the estimated optimal transition probability matrix $P_{Y|X}(y|x)$ is illustrated by Fig. 2.

The optimal transition probability under square error distortion satisfies NCE property and the result in Fig. 2 agrees with the one obtained using (6) with the estimated $P_Y(y)$, and thus the OTPM can be estimated with the previous method in [16]. In fact, as shown in Fig. 2, we can translate the NCE property into a character of the OTPM as follows.

Property 1 (Non-Crossing-Edges (NCE) property): Given a transition probability matrix $P_{Y|X}(y|x) \in \mathbb{R}^{M \times N}$ satisfying the NCE property described in Section II-A, assume the index of the first nonzero element in i-th row is f_i, and the index of the last nonzero element in $(i-1)$-th row is l_{i-1}, then $f_i \geq l_{i-1}$ for $1 \leq i \leq M - 1$.

X \ Y	0	1	2	3	4	5	6	7
0	1	0	0	0	0	0	0	0
1	0.180956	0.59298	0.226064	0	0	0	0	0
2	0	0	0.432339	0.567661	0	0	0	0
3	0	0	0	0.202804	0.797196	0	0	0
4	0	0	0	0	0.363357	0.636643	0	0
5	0	0	0	0	0	0.369901	0.630099	0
6	0	0	0	0	0	0	0.485747	0.514253
7	0	0	0	0	0	0	0	1

Fig. 2. The estimated OTPM for Example 1.

X \ Y	0	1	2	3	4	5	6	7
0	1	0	0	0	0	0	0	0
1	0.081953	0.65917	0	0.110823	0	0.111112	0	0.036942
2	0	0	0.766864	0	0	0	0.233136	0
3	0	0	0	0.91339	0	0	0	0.08661
4	0	0	0	0	1	0	0	0
5	0	0	0	0	0	0.912524	0	0.087476
6	0	0	0	0	0	0	1	0
7	0	0	0	0	0	0	0	1

Fig. 3. The estimated OTPM for Example 2.

In contrast to Example 1, in the next example we give an OTPM that does not satisfy the NCE property.

Example 2: Taking the same distribution P_X as in Example 1, we now consider the Hamming distortion metric $D_h(x, y)$ which is defined by the Hamming distance between the binary representations of x and y. For example, $D_h(0, 4) = D_h(000, 001) = 1$. Using the two-step strategy to embed 0.2 bpp message also, the estimated optimal transition probability matrix $P_{Y|X}(y|x)$ is illustrated by Fig. 3.

Observing the matrix $P_{Y|X}(y|x)$ in Fig. 3, the index of the first nonzero element in 3-th row is 3, while the index of the last nonzero element in 2-th row is 6. This means the OTPM no longer meets the Property 1, so the OTPM has to be estimated with the two-stage strategy proposed in the present paper.

For the dual problem to maximize the embedding rate subjecting to an average distortion constraint D_{av} discussed in [16], the two-stage strategy can also be adjusted to estimate the optimal $P_{Y|X}(y|x)$ for the generate distortion metric $D(x, y)$.

With this new method, we can estimate rate-distortion bound of RDH and apply RCC to approach the bound despite the distortion metrics violating NCE property. As an example, we apply the RCC to the hamming distortion case to evaluate its performance.

Example 3: We draw a 10^7-length 8-bit cover sequence $\mathbf{x} = (x_1, x_2, \cdots)$ from a discrete Laplacian distribution with mean $\mu = 0$ and scale parameter $\delta = 4$. We drop the elements whose absolute magnitude is larger than 20. The hamming distortion metric here $D_h(x, y)$ is the same as the one described in Example 2. We use the two-stage strategy to estimate the OTPM and then use RCC to embed different sizes of messages. As shown in Fig. 4, the rate distortion results are quite close to the theoretic upper bound.

The estimated optimal transition probability matrix $P_{Y|X}(y|x)$ for embedding 0.16 bpp message to the

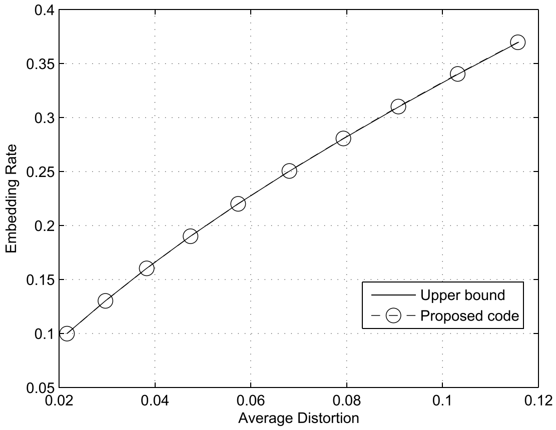


Fig. 4. Experiments on the performance of RCC for Example 3.

X \ Y	0	1	2	3	4	5	6	7	8	9
0	1	0	0	0	0	0	0	0	0	0
1	0	1	0	0	0	0	0	0	0	0
2	0	0	1	0	0	0	0	0	0	0
3	0	0	0	1	0	0	0	0	0	0
4	0	0	0	0	1	0	0	0	0	0
5	0	0	0	0	0	1	0	0	0	0
6	0	0	0	0	0	0	1	0	0	0
7	0	0	0	0	0	0	0	1	0	0
8	0	0	0	0	0	0	0	0	1	0
9	0	0	0	0	0	0	0	0	0	1
10	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0
16	0.071084	0	0	0	0	0	0	0	0	0
17	0	0.070962	0	0	0	0	0	0	0	0
18	0	0	0.071067	0	0	0	0	0	0	0
19	0	0	0	0.071063	0	0	0	0	0	0
20	0	0	0	0	0.070222	0	0	0	0	0
21	0	0	0	0	0	0.060158	0	0	0	0

Fig. 5. Part of the estimated OTPM for Example 3.

Laplacian cover sequence is partial presented in Fig. 5, from which we can see that the OPTM also violates the NCE property.

IV. Improving RCC by Compressing Cover's Histogram

For the RCC method, the cover signal's histogram $Hist_X = \{nP_X(x), x \in \mathcal{X}\}$ needs to be conveyed to the receiver, and here n is the length of the cover signal. At the sender side, the compressed $Hist_X$ will be embedded into some reserved host signals with LSB replacement, and the LSBs of these reserved signals will be reversibly embedded into other signals as part of the payload with the RCC method. At the receiver side, after recovering the probability distribution $P_X = \{P_X(x), x \in \mathcal{X}\}$ from the decompressed $Hist_X$, the recipient can calculate the optimal $P_{Y|X}(y|x)$, with which she or he can extract the message and restore the cover.

For some RDH methods, the overhead caused by conveying $Hist_X$ will greatly influence the performance of RCC. For example, in the 2D histogram based method [21], $Hist_X$ consists of bins more than 1D histogram, which will cost many pure payloads. Therefore, we should elaborate a method to compress $Hist_X$. To compress $Hist_X$, we use the differential pulse-code modulation (DPCM) encoder. Considering the Laplacian distributed histogram $Hist_X = (h_1, h_2, \ldots, h_B)$ in Fig. 6, we denote the index corresponding to the maximum value in $Hist_X$ by $iMax$, and generate the difference value sequence $\mathbf{d} = (d_1, d_2, \ldots, d_B)$ by:

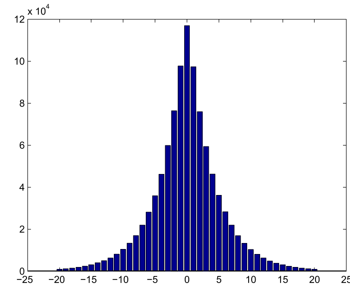


Fig. 6. Laplacian distributed frequency histogram.

$$d_i = \begin{cases} h_i - h_{i-1}, & i > 1 \text{ and } i \le iMax \\ h_i - h_{i+1}, & i < B \text{ and } i > iMax \\ h_i, & i = 1 \text{ or } i = B \end{cases} \tag{15}$$

After that, we use a M-bit sequence to record the absolute value of $\mathbf{d}$, followed by an entropy encoder to further compress the bit sequence. The parameter M is determined by the least bits to represent the maximum absolute value in $\mathbf{d}$. Meanwhile, a B-bit sequence $\mathbf{f} = (f_1, f_2, \ldots, f_B)$ is needed to record the sign of the elements in $\mathbf{d}$, which is defined as

$$f_i = \begin{cases} 0, & d_i \ge 0 \\ 1, & else. \end{cases} \quad 1 \le i \le B. \tag{16}$$

Similarly, the B-bit sequence $\mathbf{f}$ is further processed by an entropy encoder.

The total bit length of the above DPCM scheme is mainly determined by the parameter M and the histogram length B. For some irregular histograms, the maximum absolute value of $\mathbf{d}$ may be very large, so the number of bits M to represent it is big. In such cases, in order to cut the magnitude of $Hist_X$ and maintain the shape of P_X, the following quantization formula can be used to preprocess the histogram $Hist_X$:

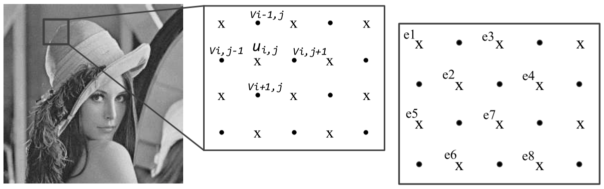
$$Hist_X = \left\lceil \frac{Hist_X}{qScale} \right\rceil, \quad P_X = \frac{Hist_X}{n/qScale} \tag{17}$$

The parameter $qScale$ here is a scale factor with common values such as 2, 4 or 8.

V. Applications Under Different Distortion Metrics

A. RDH Using Pairwise Prediction-Error Expansion

In this subsection, we improve Sachnev et al.'s method [10] and Ou et al.'s method [21] with RCC. Sachnev et al.'s method is a typical prediction-error expansion (PEE) method which generates a 1D prediction-error sequence as the cover and takes square error as the distortion metric, so its OTPM can be estimated with the previous method in [16]. Unlike conventional PEE methods, Ou et al.'s method pairs up two adjacent prediction-errors jointly to generate a sequence consisting of prediction-error pairs, and thus embeds message by modifying a 2D histogram of the prediction-errors. In Ou et al.'s method, Euclidean distance is used as the distortion metric. We will



(a) (b)

Fig. 7. Prediction pattern. (a) The pixel value u of the Cross set can be predicted by using the four neighboring pixel values of the Dot set. (b) The pairing up order for the prediction-errors.

illustrate that Euclidean distance dose not satisfy the NCE Property, and thus we have to estimate OTPM with the two-step strategy. We denote Sachnev et al.'s method as "1D Sachnev", the improved counterpart as "1D Proposed", and denote Ou et al.'s method as "2D Ou", and the improved counterpart as "2D Proposed".

Next we elaborate how to apply the RCC method to "2D Ou". In "2D Ou", the rhombus prediction pattern in Sachnev et al.'s method [10] is used to generate the prediction-errors, in which the cover image is divided into two sets denoted as "Cross" and "Dot" as shown in Fig. 7(a). The double-layered embedding scheme in Sachnev et al.'s method [10] is also applied to cover the whole image. Since the two layers' embedding processes are similar, we only take the Cross layer for illustration.

As shown in Fig. 7(b), the Cross pixels $u_{i,j}$ in the cover image are collected into a sequence $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ from left to right and from top to bottom. For each Cross pixel $u_{i,j}$, the rhombus predicted value $u'_{i,j}$ is computed using its four nearest Dot pixels:

$$u'_{i,j} = \lfloor \frac{v_{i,j-1} + v_{i+1,j} + v_{i,j+1} + v_{i-1,j}}{4} \rfloor \tag{18}$$

Then, by subtracting the predicted value $u'_{i,j}$ from the original pixel value $u_{i,j}$, we obtain the prediction-error sequence $\mathbf{e} = (e_1, e_2, \ldots, e_n)$. In an order depicted in Fig. 7(b), the prediction-error sequence is paired up to a 2D sequence $\tilde{\mathbf{e}} = ((e_1, e_2), (e_3, e_4), \ldots, (e_{n-1}, e_n))$ by taking $\tilde{e}_i = (\tilde{e}_{i1}, \tilde{e}_{i2}) = (e_{2i}, e_{2i-1})$. Here we assume the cover length n is an even number.

To extend RCC to the 2D case, we project the 2D sequence $\tilde{\mathbf{e}}$ into the 1D space using an injective mapping function *Project()* defined by

$$\begin{aligned} Project(\tilde{e}_i) &= Project(\tilde{e}_{i1}, \tilde{e}_{i2}) \\ &= (\tilde{e}_{i1} + T) \times (2 \times T + 1) + (\tilde{e}_{i2} + T) \end{aligned} \tag{19}$$

A threshold parameter T is used to choose prediction-error pairs where $|\tilde{e}_{i1}| \le T \ and \ |\tilde{e}_{i2}| \le T$. Taking the test image lena in Fig. 7(a) for instance, we set the threshold parameter $T = 7$. The 2D prediction-error histogram and the projected 1D histogram using (19) are depicted in Fig. 8(a) and Fig. 8(b) respectively.

Both Sachnev et al.'s method and the 2D Ou method use the sorting technique to pick prediction-errors with small magnitude, for the sake of generating a sharper prediction-error histogram. To be Specific, for Sachnev et al.'s 1D case, as shown in Fig. 7(a), the local variance LV for each pixel $u_{i,j}$ can be computed from the neighboring pixels $v_{i,j-1}$, $v_{i+1,j}$, $v_{i,j+1}$, and $v_{i-1,j}$ as follows:

$$LV(u_{i,j}) = \frac{1}{4}\sum_{k=1}^{4}(\Delta v_k - \Delta \bar{v}_k), \tag{20}$$

where $\Delta v_1 = |v_{i,j-1} - v_{i-1,j}|$, $\Delta v_2 = |v_{i-1,j} - v_{i,j+1}|$, $\Delta v_3 = |v_{i,j+1} - v_{i+1,j}|$, $\Delta v_4 = |v_{i+1,j} - v_{i,j-1}|$, and $\Delta \bar{v}_k = (v_1 + v_2 + v_3 + v_4)/4$.

For Ou's 2D case as shown in Fig. 9, the local variance of a pair of prediction-error $\tilde{e}_i = (e_{2i}, e_{2i-1})$ is computed from the eight neighboring pixels $p_1, \ldots, p_8$ as follows:

$$\begin{aligned} LV(\tilde{e}_i) = &|p_1 - p_2| + |p_2 - p_3| + |p_3 - p_4| \\ &+ |p_4 - p_1| + |p_3 - p_6| + |p_6 - p_5| \\ &+ |p_5 - p_4| + |p_4 - p_7| + |p_3 - p_8|. \end{aligned} \tag{21}$$

Usually we only pick out prediction-error candidates whose LV values are less than a predefined threshold T_{LV}. Considering the 2D prediction-error histogram for the test image lena in Fig. 7(a), when the sorting technique (21) is involved with threshold $T_{LV} = 20$, a sharper projected histogram is gained as shown in Fig.8(c). The sorting technique benefits RCC, because a sharper histogram has a smaller entropy so that better rate distortion performance is gained. On the other hand, the sorting technique shortens the bit length of the compressed prediction-error histogram since the number of prediction-errors is decreased.

Using (19), we project the 2D prediction-error sequence $\tilde{\mathbf{e}} = (\tilde{e}_1, \tilde{e}_2, \cdots)$ into a 1D prediction-error sequence $\mathbf{e}' = (e'_1, e'_2, \cdots)$, where $e'_i = Project(\tilde{e}_i)$. Taking $\mathbf{e}'$ as the cover sequence, we can estimate the OTPM for a given embedding rate by using the two-stage scheme described in Section III. The distortion metric $D(e'_i, e'_j)$ here is defined by the Euclidean distance between the corresponding prediction-error pairs $\tilde{e}_i$ and $\tilde{e}_j$:

$$D(e'_i, e'_j) = D(\tilde{e}_i, \tilde{e}_j) = (\tilde{e}_{i1} - \tilde{e}_{j1})^2 + (\tilde{e}_{i2} - \tilde{e}_{j2})^2 \tag{22}$$

where $e'_i = Project(\tilde{e}_i)$ and $e'_j = Project(\tilde{e}_j)$. According to the OTPM, we can reversibly embedding messages into the 2D prediction-error sequence with RCC.

When RCC is extended to the 2D case, we can see from Fig. 8(b) that the number of bins of the histogram grows a lot, and as a result the bit length to compress the cover signal's histogram $Hist_X$ becomes large. To cut the magnitude of $Hist_X$ and maintain the shape of P_X, we use the quantization scheme described in Section IV with parameter $qScale = 4$.

We compare this extended 2D RCC with three other methods, the 2D Ou method [21], our 1D RCC [19] and Sachnev et al.'s method [10]. Note that for both the 1D RCC and the extended 2D RCC, sorting techniques are added to preprocess the prediction-error sequence. While in our original 1D RCC [19], the sorting technique is not involved.

In our experiments, four test images of size 512×512 are used as covers (Fig. 10). In the 2D RCC case, for the test image Fig. 10(a), the estimated OTPM is depicted in Fig. 11 at embedding rate $r = 0.12$ bpp with $T_{LV} = 20$ and $T = 7$.

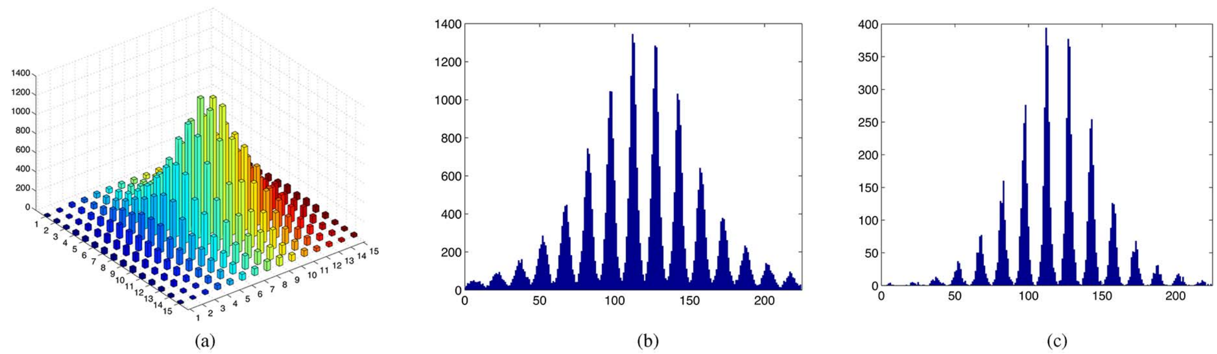


Fig. 8. Prediction-error histograms. (a) 2D prediction-error histogram. (b) Projected 1D prediction-error histogram. (c) Projected 1D prediction-error histogram with sorting technique.

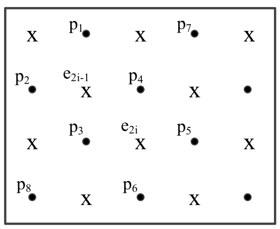


Fig. 9. Prediction context for a pair of prediction-errors.

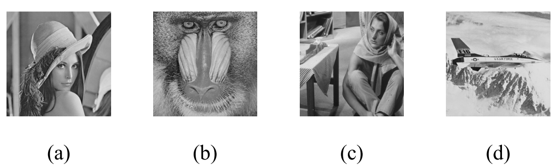


Fig. 10. Tested images of size 512×512. (a) Lena. (b) Baboon. (c) Barbara. (d) Airplane.

	Y	0	1	2	3	4	5	6	7	8	9	10	11
	0	1	0	0	0	0	0	0	0	0	0	0	0
	1	0	1	0	0	0	0	0	0	0	0	0	0
	2	0	0	1	0	0	0	0	0	0	0	0	0
	3	0	0	0.073	0.927	0	0	0	0	0	0	0	0
	4	0	0	0	0.32	0.68	0	0	0	0	0	0	0
	5	0	0	0	0	0.275	0.725	0	0	0	0	0	0
	6	0	0	0	0	0	0	1	0	0	0	0	0
	7	0	0	0	0	0	0	0	1	0	0	0	0
	8	0	0	0	0	0	0	0	0	0.477	0.523	0	0
	9	0	0	0	0	0	0	0	0	0	1	0	0
	10	0	0	0	0	0	0	0	0	0	0	0.656	0.344
X	11	0	0	0	0	0	0	0	0	0	0	0	1
	12	0	0	0	0	0	0	0	0	0	0	0	0
	13	0	0	0	0	0	0	0	0	0	0	0	0
	14	0	0	0	0	0	0	0	0	0	0	0	0
	15	0	0	0	0	0	0	0	0	0	0	0	0
	16	0	0	0	0	0	0	0	0	0	0	0	0
	17	0	0	0.102	0	0	0	0	0	0	0	0	0
	18	0	0	0.047	0.793	0	0	0	0	0	0	0	0
	19	0	0	0	0.274	0.54	0	0	0	0	0	0	0
	20	0	0	0	0	0.012	0.029	0	0	0	0	0	0
	21	0	0	0	0	0	0	0.003	0	0	0	0	0
	22	0	0	0	0	0	0	0	1	0	0	0	0
	23	0	0	0	0	0	0	0	0	0.789	0.211	0	0

Fig. 11. The estimated OTPM for 2D recursive code construction case.

As shown in Fig. 11, the NCE Property (Property 1) is obviously violated, since here the distortion metric $D(x, y)$ is abnormally user defined, which invalidates the methods in [15] and [16], and we have to estimate the OTPM with the two-step strategy.

Comparison results for various cover images are exhibited in Fig.12. From the comparison results in Fig. 12(a), Fig. 12(c) and Fig. 12(d), we can see that the 2D RCC performs the best among all the competitors. Results in Fig. 12(b) demonstrate that for high embedding rates the 1D methods can outperform the 2D methods. The reason is that the image textures in the Babbon image are more complex, thus the pixel-pairs are less correlated which decreases the performance of the 2D methods.

On the other hand, the amplitude threshold T in (19) and the sorting threshold T_{LV} in (20) and (21) can be tuned to achieve different results. Given the embedding rate $r = H(Y) - H(X)$, how to find the best values for these threshold parameters in the proposed method remains a problem to be researched in our subsequent work.

B. RDH Using Pattern Substitution in Binary Images

In [22], Ho et al. proposed a high-capacity reversible data hiding scheme in binary images based on pattern substitution (PS). The PS scheme gathers statistical data concerning the occurrence frequencies of various patterns and quantifies the occurrence frequency as it differs from pattern to pattern. In this way, some pattern exchanging relationships can be established, and pattern substitution can thus be used for data hiding. In the extraction stage, they reverse these patterns to their original forms and rebuild an undistorted cover image.

Given the original binary image H, the PS method uses the image differencing D to hide messages. Image differencing is a method used to identify the edges of a binary image, so the changes to D results in modifying the pixels on the edge of the original cover image H, which is not discernible to human eyes. The image differencing D is defined by

$$D(i,j) = \begin{cases} H(i,j) & \text{if } i = 1 \textit{ and } j = 1 \\ H(i,j) \oplus H(i-1,j) & \text{if } i \neq 1 \text{ and } j = 1, \\ H(i,j) \oplus H(i,j-1) & \text{otherwise}, \end{cases} \tag{23}$$

where $\oplus$ represents the Exclusive-OR logic operation, and i and j represent the rows and columns of the cover image respectively. D represents the image differencing and H is the original binary image.

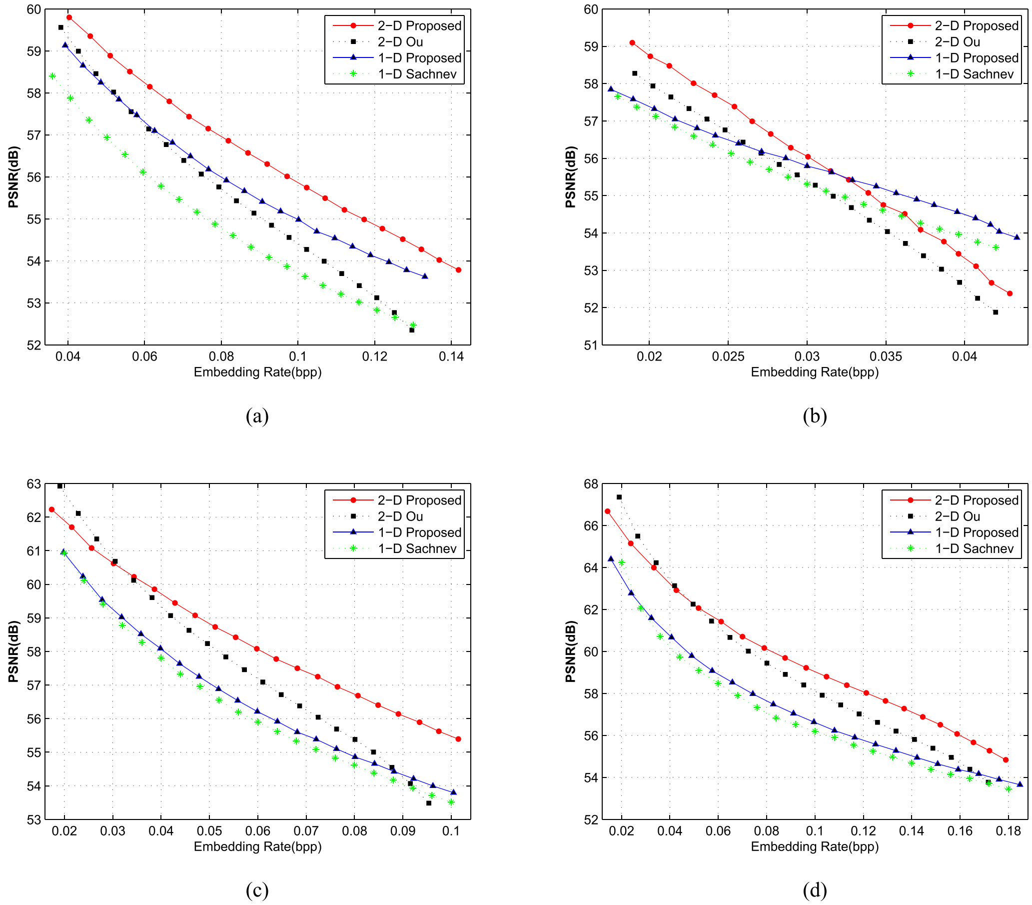
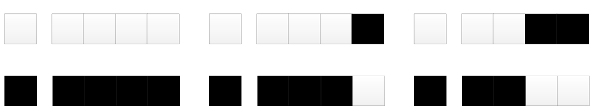
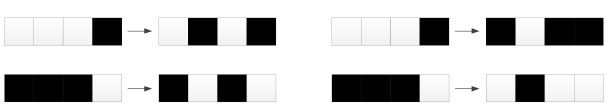


Fig. 12. Comparisons in terms of rate-distortion performance. (a) Lena. (b) Baboon. (c) Barbara. (d) Airplane.



(a) (b) (c)

Fig. 13. The results of the patterns against the original image pixels in H. (a) P_{0000}. (b) P_{0001}. (c) P_{0010}.

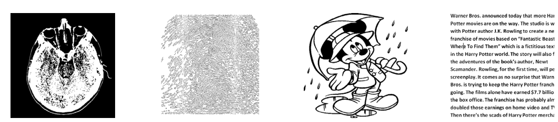


(a) (b)

Fig. 14. Pattern substitution against the original pixel modifications in H. (a) P_{0001} to P_{0111}. (b) P_{0001} to P_{1110}.

After the image differencing, the PS method picks a large number of continuous sets of 4 pixels from D to form patterns as depicts in Fig.13. Assume that, for binary images, "1" stands for the black pixels, and "0" stands for the white pixels. Sixteen patterns exist, and among them, if the number of "1" is greater, the occurrence probability is lower. This is because "1" in image D only occurs where the black and white pixels interlace in the cover image H.

The other characteristic is that if only one bit in D is modified, all the pixels of the rebuilt image following the



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

Fig. 15. Test binary images of size 512 × 512. (a) CT. (b) Finger print. (c) Mickey. (d) English text.

Y

X	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.984	0	0	0	0	0	0	0	0	0	0	0	0.01578	0	0
2	0	0.98	0	0	0	0	0	0	0	0	0	0	0	0.02	0
3	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0.98	0	0	0.02	0	0	0	0	0	0	0	0
5	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	1	0	0	0	9.36E-06	0	0	0	0
8	0	0	0	0	0	0	0	0.983	0	0	0.01669	0	0	0	0
9	0	0	0	0	0	0	0	0	0.98	0	0	0	0	0	0.02
10	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
13	0	0	0	0	0	0	0	0	0	0	1.60E-06	0	1	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	6.80E-06	1	0
15	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Fig. 16. The estimated OTPM for embedding 3500 bits into the binary English text image.

modification will be different from their originals in the initial image H. In order to prevent serious damage to the original image H, all patterns in image D are divided into two groups according to the number of "1", named "the odd group" and "the even group", and pattern substitutions across groups are

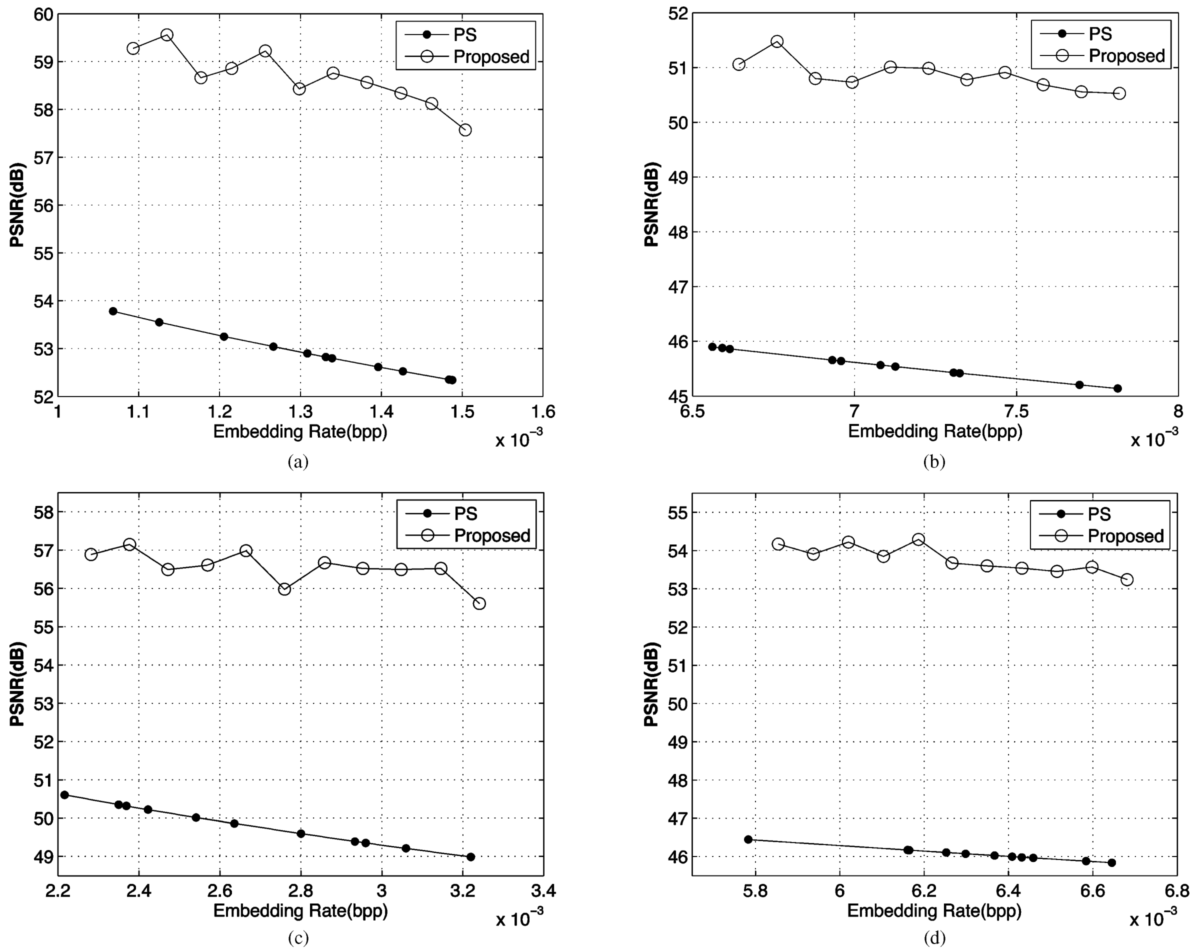


Fig. 17. Comparisons of the PS method and the proposed method. (a) CT. (b) Finger print. (c) Mickey. (d) English text.

not allowed. In other words, if a pattern has an odd number of "1" bits, the number of "1" bits must remain odd after the modification; likewise, if a pattern has an even number of "1" bits, the number of "1" bits also has to remain even after modification.

Inside each group, substitutions of its patterns in D will make modification to the original image pixels in H. Taking the pattern P_{0001} in the odd group for instance, as shown in Fig. 14(a), we find that if P_{0001} is replaced by P_{0111}, only one bit in the cover image H will be modified. According to Fig. 14(b), if P_{0001} is replaced by P_{1110}, three bits in H will be modified.

From the above observations, for each pattern PM in one group, the PS method can always pick out another pattern PF in the same group that is modified the least and then pair them up to hide messages. When hiding "0", the PM pattern stays unchanged, when hiding "1", the PM pattern is changed to PF. To get a high embedding capacity, the PM pattern should have a high occurrence frequency and the PF pattern should have a low occurrence frequency. After choosing a pair of patterns, the PS scheme scans the image D from left to right and from top to bottom to record the positions of PM and PF to hide messages. We refer readers to [22] for the details of the PS method.

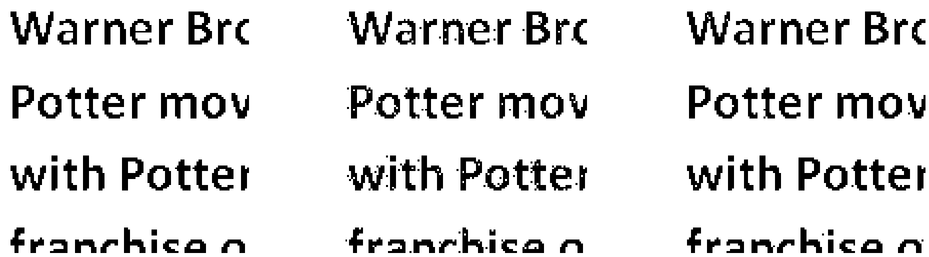
To extend RCC method to this case, we present the image differencing D as a 4D sequence $\mathbf{d} = (d_1, d_2, \ldots, d_n)$ where each element $d_i = (d_{i1}d_{i2}d_{i3}d_{i4})$ is a binary 4-tuple. For instance, the tuples (0000), (0001) and (0010) stands for the patterns in Fig. 13. For each 4-tuple d_i, the corresponding 4-tuple in the cover image H is denoted as $h_i = (h_{i1}h_{i2}h_{i3}h_{i4})$.

Since the elements of the 4-tuples $\{d_i, i = 1, 2, \ldots, n\}$ only take binary value, we can cast them into a decimal number and get a mapped 1D sequence $\tilde{\mathbf{d}} = (\tilde{d}_1, \tilde{d}_2, \ldots, \tilde{d}_n)$ while $\tilde{d}_i \in \{0, 1, 2, \ldots, 15\}$. We define the distortion metric $D(\tilde{d}_i, \tilde{d}_j)$ by

$$D(\tilde{d}_i, \tilde{d}_j) = \begin{cases} HamDist(h_i, h_j), & Group(d_i) = Group(d_j) \\ +Inf, & otherwise \end{cases} \tag{24}$$

where *HamDist()* above denotes the hamming distance for two binary tuples. The value *Group(d_i)* will be 0 if the number of bit "1" in d_i is even, otherwise will be 1. Pattern substitutions across groups are not allowed as discussed before.

After that, we apply the RCC to the decimal sequence $\tilde{\mathbf{d}}$ with the specified distortion metric $D(\tilde{d}_i, \tilde{d}_j)$ and compare it with the PS method for different binary cover images. For clarity, we only consider the non-overlapping patterns in [22].



(a) (b) (c)

Fig. 18. Comparisons of visual quality for the English text image. (a) Cover image. (b) PS method. (c) Proposed method.

Four binary test images shown in Fig. 15 are used in our experiments. For the PS method, although the picked pair of patterns PM and PF only makes one bit modification to the original cover image H, it has to take extra bits to record the place of PF for all embedding capacities. This makes it less adaptable especially for low embedding rates. To the contrary, as shown by the estimated OTPM $P_{Y|X}(y|x)$ in Fig.16, the RCC method can automatically pick out different pairs of pattern transitions for a given embedding rate. Fig.16 also implies that the NCE Property no longer meets here, and thus we have to estimate the OTPM with the two-step strategy.

The PS method has an inherent limitation that a pattern can only changed to another pattern, which means it can only embed $\log_2(2) = 1$ bit for a pair. While for high embedding capacities, the RCC method can adaptively pick out transitions from one pattern to two or more patterns. For example, a triple of patterns can embed $\log_2(3) = 1.585$ bits at most. These two aspects above can explain that the RCC makes less modifications than the PS method and therefore earns better embedding performance as shown in Fig. 17.

As used in [22], the PSNR value for the binary images here is calculated as follows:

$$PSNR = 10 \log_{10} \frac{255^2}{MSE} \; dB \tag{25}$$

while MSE is computed by

$$MSE = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} (b_{ij} - b'_{ij})^2 \times 255 \tag{26}$$

where M and N are the width and height of the binary cover image, b_{ij} means the pixel value of the original binary image in position (i, j) and b'_{ij} represents the pixel value after modification in position (i, j).

To better elaborate the advantage of the proposed method, comparisons of visual quality between the two methods for the English test image Fig. 15(d) are illustrated in Fig.18. From the results we can see that the proposed method made fewer modifications to the binary test image.

VI. Conclusions and Future Work

In this paper, we propose a unified algorithm to estimate the optimal transition probability matrix $P_{Y|X}(y|x)$ for the general distortion metrics $D(x, y)$, which enables us to extend the recursive code construction to various applications. The experiments demonstrate that the proposed method outperforms state-of-art algorithms significantly.

The proposed framework for reversible data hiding shows its generalizing ability and good performance under different application scenarios. But when extended to higher-dimensional cases, 3D or 4D, the cover signal's frequency histogram contains many number of bins, so the compression of the frequency histogram may comes to a problem which needs to be better solved. On the other hand, an optimal injective transition probability matrix $P_{Y|X}(y|x)$ instead of the many-to-one transition probability matrix seems more preferable, since it can simplify the coding and decoding processes for data embedding. These aspects may be our continual topics to work with.

Acknowledgment

The authors would like to express our sincere thanks to the anonymous reviewers for their valuable comments, and thanks to B. Ou et al. for offering the codes of their paper [21].

References

[1] A. Khan, A. Siddiqa, S. Munib, and S. A. Malik, "A recent survey of reversible watermarking techniques," *Inf. Sci.*, vol. 279, pp. 251–272, Sep. 2014. [Online]. Available: http://dx.doi.org/10.1016/j.ins.2014.03.118

[2] A. S. Brar and M. Kaur, "Reversible watermarking techniques for medical images with ROI-temper detection and recovery—A survey," *Int. J. Emerg. Technol. Adv. Eng.*, vol. 2, no. 1, pp. 32–36, Jan. 2012.

[3] J. Fridrich, M. Goljan, and R. Du, "Lossless data embedding for all image formats," *Proc. SPIE*, vol. 4675, pp. 572–583, Jan. 2002.

[4] K.-L. Chung, Y.-H. Huang, P.-C. Chang, and H.-Y. M. Liao, "Reversible data hiding-based approach for intra-frame error concealment in H.264/AVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 11, pp. 1643–1647, Nov. 2010.

[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] 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.

[7] Y. Hu, H.-K. Lee, and J. Li, "DE-based reversible data hiding with improved overflow location map," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 2, pp. 250–260, Feb. 2009.

[8] 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, Mar. 2006.

[9] P. Tsai, Y.-C. Hu, and H.-L. Yeh, "Reversible image hiding scheme using predictive coding and histogram shifting," *Signal Process.*, vol. 89, no. 6, pp. 1129–1143, 2009.

[10] 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.

[11] L. Luo, Z. Chen, M. Chen, X. Zeng, and Z. Xiong, "Reversible image watermarking using interpolation technique," *IEEE Trans. Inf. Forensics 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 Process.*, 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 Process.*, vol. 20, no. 12, pp. 3524–3533, Dec. 2011.

[14] T. Kalker and F. M. J. Willems, "Capacity bounds and constructions for reversible data-hiding," in *Proc. 14th Int. Conf. Digit. Signal Process. (DSP)*, 2002, pp. 71–76.

[15] 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. Inf. Forensics Security*, vol. 7, no. 4, pp. 1155–1167, Aug. 2012.

[16] 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. Inf. Forensics Security*, vol. 8, no. 5, pp. 779–788, May 2013.

[17] W. Zhang, B. Chen, and N. Yu, "Capacity-approaching codes for reversible data hiding," in *Proc. 13th Inf. Hiding Conf.*, May 2011, pp. 255–269.

[18] W. Zhang, B. Chen, and N. Yu, "Improving various reversible data hiding schemes via optimal codes for binary covers," *IEEE Trans. Image Process.*, vol. 21, no. 6, pp. 2991–3003, Jun. 2012.

[19] 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, Jul. 2013.

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

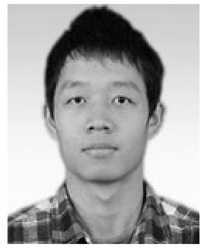
[21] B. Ou, X. Li, Y. Zhao, R. Ni, and Y.-Q. Shi, "Pairwise prediction-error expansion for efficient reversible data hiding," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 5010–5021, Dec. 2013.

[22] Y.-A. Ho, Y.-K. Chan, H.-C. Wu, and Y.-P. Chu, "High-capacity reversible data hiding in binary images using pattern substitution," *Comput. Standards Inter.*, vol. 31, no. 4, pp. 787–794, Jun. 2009.

[23] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge Univ. Press, 2004.

[24] F. Willems, D. Maas, and T. Kalker, "Semantic lossless source coding," in *Proc. 42nd Annu. Allerton Conf. Commun., Control Comput.*, Sep. 2004, pp. 1–8.

Weiming Zhang received the M.S. and Ph.D. degrees from the Zhengzhou Information Science and Technology Institute, Zhengzhou, China, in 2002 and 2005, respectively. He is currently an Associate Professor with the School of Information Science and Technology, University of Science and Technology of China, Hefei, China. His research interests include multimedia security, information hiding, and cryptography.



Xiaocheng Hu received the B.S. degree from the University of Science and Technology of China, Hefei, China, in 2010, where he is currently pursuing the Ph.D. degree. His research interests include multimedia security, image and video processing, video compression, and information hiding.

Xiaolong Li received the B.S. degree from Peking University, Beijing, China, in 1999, the M.S. degree from École Polytechnique, Palaiseau, France, in 2002, and the Ph.D. degree in mathematics from the École Normale Supérieure de Cachan, Cachan, France, in 2006. He was a Post-Doctoral Fellow with Peking University from 2007 to 2009, where he is currently a Researcher. His research interests are image processing and information hiding.

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