Multimed Tools Appl
DOI 10.1007/s11042-016-3850-z

Optimal structural similarity constraint for reversible data hiding

Jiajia Xu[1] · Weiming Zhang[1] · Ruiqi Jiang[1] · Xiaocheng Hu[1] · Nenghai Yu[1]

Received: 11 March 2016 / Revised: 18 July 2016 / Accepted: 8 August 2016

Abstract Until now, most reversible data hiding techniques have been evaluated by peak signal-to-noise ratio(PSNR), which based on mean squared error(MSE). Unfortunately, MSE turns out to be an extremely poor measure when the purpose is to predict perceived signal fidelity or quality. The structural similarity (SSIM) index has gained widespread popularity as an alternative motivating principle for the design of image quality measures. How to utilize the characterize of SSIM to design RDH algorithm is very critical. In this paper, we propose an optimal RDH algorithm under structural similarity constraint. Firstly, we deduce the metric of the structural similarity constraint, and further we prove it does't hold non-crossing-edges property. Secondly, we construct the rate-distortion function of optimal structural similarity constraint, which is equivalent to minimize the average distortion for a given embedding rate, and then we can obtain the optimal transition probability matrix under the structural similarity constraint. Comparing with previous RDH, our method have

This work was supported in part by the Natural Science Foundation of China under Grants 61170234 and 60803155, and by the Strategic Priority Research Program of the Chinese Academy of Sciences under Grant XDA06030601.

✉ Weiming Zhang
zhangwm@ustc.edu.cn

Jiajia Xu
xujiajia@mail.ustc.edu.cn

Ruiqi Jiang
rqjiang@mail.ustc.edu.cn

Xiaocheng Hu
hxc@mail.ustc.edu.cn

Nenghai Yu
ynh@mail.ustc.edu.cn

[1] School of Information Science and Technology, University of Science and Technology of China, Hefei 230026, China

gained the improvement of SSIM about 1.89 % on average. Experiments show that our proposed method outperforms the state-of-arts performance in SSIM.

Keywords Reversible data hiding · Structural similarity · Recursive code construction · Convex optimization

1 Introduction

Reversible data hiding (RDH) [3, 12, 15, 16, 20] as a special branch of information hiding, it is not only concerned about the users embedding data, but also pay attention to the carriers themselves. It requires the carriers to be completely recovered after extracting the embedded message, which has been found to be useful in many fields, such as, medical imagery and legal. In the past few years, reversible hiding has been considerably developed. Scholars have proposed a variety of reversible hiding algorithms for digital images, digital videos, audios, and other carriers. Most RDH algorithms consist three key steps. The first step is predicting, which focuses on how to better exploit inter-pixel correlations to derive a sharply distributed one. The second step is sorting technique, which exploits the correlation between neighboring pixels for optimizing embedding order. The third step reversibly embeds the message into the prediction-error by modifying its histogram.

Until now, in order to facilitate the efficiency of RDH, researchers have proposed many methods in the past decade. In general, RDH algorithms roughly fall into three categories: the compression appending framework [3], the histogram shift (HS) technique [12] and the difference expansion (DE) scheme [20]. In [3], Fridrich et al. proposed to find the space by compressing proper bit-plane with the minimum redundancy. In their method, unless the image is noisy, the lowest bit-plane is compressed and embedded with a hash value. However, the above [3] method cannot yield a satisfactory performance, since the correlations among a bit-plane is too weak to provide a high embedding capacity. HS technique is first proposed by Ni et al. [12] and this type of schemes are implemented by modifying the image histogram of a certain dimension. In [20], Tian introduced a DE technique, which discovers extra storage space by exploring the redundancy in the image content. He employ the DE technique to reversibly embed a payload into images. The DE can achieve high embedding capacity and keep the distortion low, while comparing with the lossless-compression-based schemes [3] and HS-based scheme [12], the DE method performs much better by providing a higher embedding capacity while keeping the distortion low. Unlike in DE where only the correlation of two adjacent pixels is considered, the local correlation of larger neighborhood is exploited in prediction-error expansion (PEE) [21], and thus a better performance can be expected. PEE is currently a research hot spot and the most powerful technique of RDH.

Almost RDH techniques have been evaluated by PSNR. The PSNR is based on MSE. What is the MSE? The definition of MSE between $\mathbf{x} = (x_1, x_2, \cdots, x_N)$ and $\mathbf{y} = (y_1, y_2, \cdots, y_N)$ is $MSE(\mathbf{x}, \mathbf{y}) = 1/N \sum_{i=1}^{N}(x_i - y_i)^2$, and PSNR is $PSNR = 10log_{10}(L^2/MSE)$. where L is the dynamic range of allowable image pixel intensities. The MSE has many attractive features. Firstly, the MSE is very simple, and satisfy properties of convexity, symmetry, and triangular inequality. All L_p norms are excellent distance metrics in N-dimensional Euclidean space, especially in the context of optimization. Secondly, the MSE has a clear physical meaning, which is the natural way to define the energy of the error signal. Thirdly, the MSE is a desirable measure in the statistics and estimation framework.The MSE has become a convention in many applications. Unfortunately, MSE turns out to be an poor measure when the purpose is to predict perceived signal fidelity or quality

[20, 21]. There are several implicit assumptions when using MSE, such as signal fidelity is independent of temporal or spatial relationships among the original signal, the error signal and the samples of the original signal, and is independent of the signs of the error signal samples. However, It is a pity that not one of assumptions hold when we using MSE to measure the visual perception of image fidelity. The other reasons lead MSE to be poor measure can be found in paper [23].

Due to the limitations or poor performance of MSE, what is the alternative? Recently, the SSIM as novel image fidelity or similarity measures, which was originally motivated by the observation that natural image signals are highly structured, has attracted a great deal of attention [1, 23]. The human visual system (HVS) is the principle philosophy of SSIM approach, which is highly sensitive to the structural distortions and automatically compensates for the nonstructural distortions. The basic ideas of SSIM approach is simulating the HVS functionality, which prove highly effective for measuring the similarity. Therefore, the SSIM has gained widespread popularity as an alternative motivating principle for the design of image quality measures in many applications, such as image fusion, image compression, video hashing, chromatic image quality, retinal and wearable displays, and ratedistortion optimization in standard video compression [1, 23, 25] . However, how to utilize the characteristics of SSIM to design RDH algorithm is very critical. In practice, SSIM is often used as a black box in optimization tasks as merely an adhesive control unit outside the main optimization module. Brunet et al. [1] construct a series of normalized and generalized (vector-valued) metrics based on the important ingredients of SSIM, and show that such modi?ed measures are valid distance metrics and have many useful properties, such as quasi-convexity, a region of convexity around the minimizer, and distance preservation under orthogonal or unitary transformations.

In this paper, we propose an optimal structural similarity constraint for RDH algorithm by utilizing the characterize of SSIM. Firstly, $SSIM(x, y)$ is not a metric, we should design the corresponding structural similarity constraint in order to approach the upper bound of the payload. Based on Brunet et al. [1], we deduce the metric of the structural similarity constraint, and further we prove it does't hold non-crossing-edges property. In this condition, we use the Earth Movers Distance strategy in [18] to estimate the optimal transition probability matrix. Secondly, we construct the rate-distortion function of optimal structural similarity constraint, which is equivalent to minimize the average distortion for a given embedding rate, and then we can obtain the optimal transition probability matrix under the structural similarity constraint. Experiments show that our proposed method can be used to improve the performance of previous RDH schemes evaluated by SSIM, especially under high embedding rates. Both of this indicate that our proposed OSSC algorithm is obvious effect for RDH.

The paper is organized as follows. Section 2 describes the proposed optimal structural similarity constraint for reversible data hiding. The simulations done using the proposed technique and the obtained results are presented in Section 3. In Section 4, onclusions are briefly drawn based on the results.

2 Optimal structural similarity constraint (OSSC) for RDH

2.1 The fundamental of stuctural similarity index

In the paper [23], Wang et al. has proposed the SSIM for image quality assessment, which compares local patterns of pixel intensities that have been normalized for luminance, contrast and structure. Suppose that $x \in \mathbf{R}_+^N$ and $y \in \mathbf{R}_+^N$ are local image signals, which are taken from two images in the same location. The SSIM index separates the task of similarity measurement into three comparisons: luminance, contrast and structure, and the three components are relatively independent.The first is luminance similarity $l(x, y)$, which is relevant with the mean intensity u_x, u_y and qualitatively consistent with Weberslaw.

$$l(x, y) = \frac{2u_x u_y + c_1}{u_x^2 + u_y^2 + c_1} \tag{1}$$

The second is contrast similarity function $c(x, y)$, Which is relevant with the variance σ_x, σ_y and consistent with the contrast-masking feature of the HVS.

$$c(x, y) = \frac{2\sigma_x \sigma_y + c_2}{\sigma_x^2 + \sigma_y^2 + c_2} \tag{2}$$

The third is structure similarity is $s(x, y)$, Which is conducted on these normalized signals $(x - u_x)/\sigma_x$ and $(y - u_y)/\sigma_y$.

$$s(x, y) = \frac{\sigma_{x,y} + c_3}{\sigma_x \sigma_y + c_3} \tag{3}$$

where $u_x, u_y, \sigma_x, \sigma_y$, and represent, respectively, the mean and variance of x and y. $\sigma_{x,y}$ represent the covariance between x and y. The constants c_1, c_2, c_3 are included to avoid instability when $u_x^2 + u_y^2$, $\sigma_x^2 + \sigma_y^2$ and $\sigma_x \sigma_y$ are very close to zero, respectively. Then, we combine the three comparisons of luminance, contrast and structure, and get the SSIM index function.

$$SSIM(x, y) = \frac{2u_x u_y + c_1}{u_x^2 + u_y^2 + c_1} \times \frac{2\sigma_x \sigma_y + c_2}{\sigma_x^2 + \sigma_y^2 + c_2} \times \frac{\sigma_{x,y} + c_3}{\sigma_x \sigma_y + c_3} \tag{4}$$

The SSIM index is computed locally within a sliding window that moves pixel-by-pixel across the image.The boundedness of SSIM is $1 \geq |SSIM(x, y)|$. Only if $x = y$,the $SSIM(x, y) = 1$. That is to say, the closer that x and y are to each other, the closer $SSIM(x, y)$ is to 1. Besides SSIM have symmetrical $SSIM(x, y) = SSIM(y, x)$.

2.2 The metric of stuctural similarity index

Based on (4) , we should design the corresponding structural similarity constraint in order to approach the upper bound of the payload. Does the $SSIM(x, y)$ is a distortion metric ? A metric $D(x, y)$ must satisfy four rules for all $x, y, z \in \mathbf{R}_+^N$, as follows:

- nonnegativity: $D(x, y) \geq 0$.
- symmetry: $D(x, y) = D(x, y)$.
- identity:$D(x, y) = 0$ if and only if $x = y$.
- triangular inequality: $D(x, y) + D(y, z) \geq D(x, z)$.

Clearly, the SSIM index is not a metric, because $x = y \Rightarrow SSIM(x, y) = 0$ and $SSIM(x, y) + SSIM(y, z) \geq SSIM(x, z)$ is not established. However, what's the metric can be used to characterize the structural similarity constraint? We will find a way to shape it to form a metric. Based on (4), we set $c_3 = c_2/2$, then we can get $SSIM(x, y)$ as follows:

$$SSIM(x, y) = \frac{2u_x u_y + c_1}{u_x^2 + u_y^2 + c_1} \times \frac{2\sigma_{x,y} + c_2}{\sigma_x^2 + \sigma_y^2 + c_2} \tag{5}$$

Given $x, y \in \mathbf{R}$, $\mathbf{R}$ is a normed space,a normalized metric or relative distance is a metric of the form

$$d_n(x, y) = \frac{\|x - y\|}{(\|x\|^p + \|y\|^p)^{\frac{q}{p}}} \tag{6}$$

$d_n(x, y)$ is a metric for $q = 1$ and for all $p \geq 1$. Brunet et al. [1] construct a series of normalized and generalized metrics based on the important ingredients of SSIM, and show that such modified measures are valid distance metrics and have many useful properties. Considering set $q = 1$ and $p = 2$, which leads to

$$\begin{aligned} d_1(u_x, u_y) &= \sqrt{\frac{\|u_x - u_y\|^2}{(\|u_x\|^2 + \|u_y\|^2 + c_1)}} \\ &= \sqrt{1 - \frac{2u_x u_y + c_1}{u_x^2 + u_y^2 + c_1}} \end{aligned} \tag{7}$$

$$\begin{aligned} d_2(x - u_x, y - u_y) &= \sqrt{\frac{\|(x - u_x) - (y - u_y)\|^2}{(\|x - u_x\|^2 + \|y - u_y\|^2 + c_2)}} \\ &= \sqrt{\frac{\sigma_x^2 - 2\sigma_{x,y} + +\sigma_y^2}{\sigma_x^2 + \sigma_y^2 + c_2}} \\ &= \sqrt{1 - \frac{2\sigma_{x,y} + c_2}{\sigma_x^2 + \sigma_y^2 + c_2}} \end{aligned} \tag{8}$$

Observing (8), it is not difficult to find that the relationship among $SSIM(x, y)$, $d_1(u_x, u_y)$ and $d_2(x - u_x, y - u_y)$. which can be writen by

$$\frac{2u_x u_y + c_1}{u_x^2 + u_y^2 + c_1} = 1 - d_1(u_x, u_y)^2 \tag{9}$$

$$\frac{2\sigma_{x,y} + c_2}{\sigma_x^2 + \sigma_y^2 + c_2} = 1 - d_2(x - u_x, y - u_y)^2 \tag{10}$$

based on $SSIM(x, y) = (1 - d_1(u_x, u_y)^2)(1 - d_2(x - u_x, y - u_y)^2)$,we can get

$$\sqrt{1 - SSIM(x, y)} = \sqrt{1 - d_1^2 + d_2^2 - d_1^2 d_2^2} \tag{11}$$

$$\|d(x, y)\|^2 = 1 - SSIM(x, y) \tag{12}$$

If $u_x = u_y$ and $x - u_x = y - u_y$, then $SSIM(x, y) = 1$, so we can have

$$\|d(x, y)\|^2 = \frac{\|x - y\|^2 + c}{\|x\|^2 + \|y\|^2 + c} \tag{13}$$

where $c \geq 0$. Now, we should verify the $\|d(x, y)\|^2$ whether met the criteria of metric.

- Firstly, because $1 \geq |SSIM(x, y)|$ and $\|d(x, y)\|^2 = 1 - SSIM(x, y)$, so $\|d(x, y)\|^2 \geq 0$.
- Secondly, because $SSIM(x, y) = SSIM(y, x)$ and $1 - SSIM(x, y) = 1 - SSIM(y, x)$, so we can get $\|d(x, y)\|^2 = \|d(y, x)\|^2$.
- Thirdly, if and only if $x = y$, the $SSIM(x, y) = 1$ and $\|d(x, y)\|^2 = \|d(y, x)\|^2 = 0$.
- The last is triangular inequality. Refer to the method in D. Brunet's paper [1], we can prove the property of triangular inequality: $\|d(x, y)\|^2 + \|d(y, x)\|^2 \geq \|d(y, x)\|^2$.

So the $\|d(x, y)\|^2$ is a metric, which can be used to characterize the property of $SSIM(x, y)$. The $\|d(x, y)\|^2$ hold many mathematical properties such as convexity, quasi-convexity, and generalized convexity, which can be derived from SSIM.

2.3 Using earth movers distance to solve the structural similarity constraint for RDH

How to get the optimal transition probability matrix for a given embedding rate RDH under structural similarity constraint? Kalker and Willems [8] formulated the RDH as a special rate-distortion problem. For independent and identically distributed (i.i.d.) host signals, the upper bound of the payload and a distortion constraint is given by Kalker and Willems [8]. They obtained the rate-distortion function under a given distortion constraint, as follows:

$$\rho_{rev}(\Delta) = maximize\{H(Y)\} - H(X) \tag{14}$$

where X and Y denote the random variables of the host signal and the marked signal respectively. The maximum entropy is over all transition probability matrices $P_{Y|X}(y|x)$ satisfying the distortion constraint $\sum_{x,y} P_X(x) P_{Y|X}(y|x) D(x, y) \leq \Delta$. $D(x, y)$ is the distortion metric. Consequently, the optimal transition probability matrix (OTPM) $P_{Y|X}(y|x)$ for (14) implies the optimal modification manner on the histogram of the host signal X. For a binary host sequence, i.e., $x \in 0, 1$, Kalker and Willems [8] proposed a recursive code construction and Zhang et al. [15, 31] improved the recursive code construction to approach the rate-distortion bound. For some gray-scale signals and specific distortion metrics $D(x, y)$, such as square error distortion $D(x, y) = (x - y)^2$ or L_1-Norm $D_1(x, y) = |x - y|$,the OTPM has a Non-CrossingEdges property [11]. 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 $P_Y(y)$. In [7], Hu et al. proposed a fast algorithm to estimate the optimal marginal distribution $P_Y(y)$ for both the distortion constrained problem (14) and its dual problem, i.e., the embedding rate constrained problem. However, for some distortion metrics, such as Hamming distance, the NCE property no longer holds and the OTPM can not be obtained analytically.

Property 1 (Non-Crossing-Edges (NCE) property) : 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. [11] has been proved that when the distortion metrics $D(x, y) = (x - y)^2$ or $D(x, y) = |x - y|$, the transition probability matrix $P_{Y|X}(y|x)$ has the NCE property. Does $\|d(x, y)\|^2 = (\|x - y\|^2 + c)/(\|x\|^2 + \|y\|^2 + c)$ meet the criteria of NCE ?

$$\frac{\partial \|d(x, y)\|^2}{\partial x} = \frac{2(x^2 - y^2 - c)}{(x^2 + y^2 + c)^2} \tag{15}$$

$$\frac{\partial \|d(x, y)\|^2}{\partial x} \begin{cases} \geq 0, x \geq \sqrt{y^2 + c} \\ < 0, x < \sqrt{y^2 + c} \end{cases} \tag{16}$$

When the $x \geq \sqrt{y^2 + c}$, the $\partial \|d(x, y)\|^2/\partial x \geq 0 \Rightarrow \|d(x, y)\|^2$ is increasing function. On the other hand, when the $x < \sqrt{y^2 + c}$, the $\partial \|d(x, y)\|^2/\partial x < 0 \Rightarrow \|d(x, y)\|^2$ is strictly decreasing function. For any $P_{Y|X}(y_1|x_1) > 0$ and $P_{Y|X}(y_2|x_2) > 0$, if

$max(\sqrt{y_1^2+c}, \sqrt{y_2^2+c}) \leq x_1 < x_2 \Rightarrow \|d(x_1, y_1)\|^2 \leq \|d(x_2, y_1)\|^2, \|d(x_1, y_2)\|^2 \leq \|d(x_2, y_2)\|^2$. In the paper [11], Lin define a function $g(x, y) = -log_2 P_Y^{\dagger}(y) - \lambda\|d(x, y)\|^2$. If we want to prove the NEC property, only need to prove formula holds $g(x_1, y_1) + g(x_2, y_2) \geq g(x_1, y_2) + g(x_2, y_1)$. That is to say $-\lambda\|d(x_1, y_1)\|^2 - \lambda\|d(x_2, y_2)\|^2 \geq \lambda\|d(x_1, y_2)\|^2 - \lambda\|d(x_2, y_1)\|^2$, and this problem is equivalent to $\|d(x_2, y_1)\|^2 - \|d(x_1, y_1)\|^2 \geq \|d(x_2, y_2)\|^2 - \|d(x_1, y_2)\|^2$, which does not always holds. So the $\|d(x, y)\|^2$ does not meet the criteria of NCE. If the NCE property no longer hold and then the optimal transition probability matrix cannot be obtained analytically. However, how to efficiently solve the problem, realizing the optimal modification for gray-scale signals, i.e., $x \in \{0, 1, \cdots, B-1\}$ remains a problem

Fortunately, the Earth Movers Distance (EMD) proposed by Rubner et al. [18] is defined as the minimal cost that must be paid to transform one histogram into the other, where there is a ground distance between the basic features that are aggregated into the histogram. Figuratively speaking, the EMD gains its name from the intuition that given two distributions, one can be seen as a mass of earth properly spread in space, the other as a collection of holes in that same space [5]. Then, the EMD measures the least amount of work needed to fill the holes with earth, where a unit of work corresponds to transporting a unit of earth by a unit of ground distance. The EMD has many advantages over other similarity measures for distributions [22]. Firstly, the EMD matches perceptual similarity better than bin-bybin distances for histogram matching. Secondly, the cost of moving "earth" reflects the notion of nearness properly, without the quantization problems of most current measures. Thirdly, computing the EMD is based on a solution to the wellknown transportation problem from linear optimization, for which efficient algorithms, e.g., simplex methods, are available [2] (Fig. 1).

In this paper, we formulate the EMD in the specific context of RDH, where the EMD is employed to optimal transition probability matrices $P_{Y|X}(y|x)$. Assume that a memoryless source produces the host sequence $\mathbf{x} = (x_1, x_2, \cdots, x_N)$ with the identical distribution $P_X(x)$ such that $x \in \{0, 1, \cdots, B-1\}$, where $B \geq 1$ is an integer. The message is usually encrypted before being embedded, so we assume that the secret message $\mathbf{m} = (m_1, m_2, \cdots)$ is a binary random sequence with $m_i \in \{0, 1\}$. Through slightly modifying its elements to produce the marked-sequence $\mathbf{y} = (y_1, y_2, \cdots, y_N)$. Based on the property of $SSIM(x, y)$, we selected $\|d(x, y)\|^2$ as structural similarity distortion constraint. $R = L/n$ is embedding rate in RDH. Based on $\rho_{rev}(\Delta) = maximize\{H(Y)\} - H(X)$, the mathematical model can

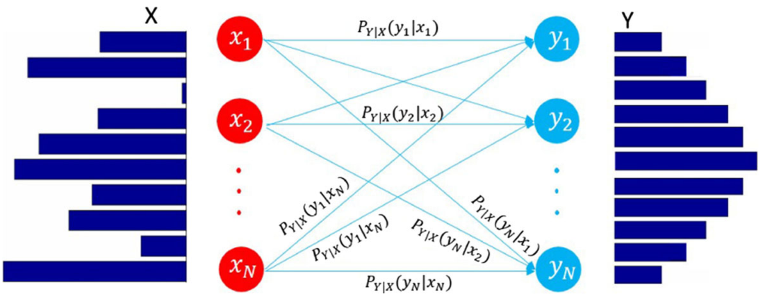


Fig. 1 EMD between two equal signatures as a transportation problem

be equivalent to minimize the average distortion for a given embedding rate R , which is formulated as:

$$EMD(X,Y) = \frac{\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} P_{Y|X}(y|x)\|d(x,y)\|^2}{\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} P_{Y|X}(y|x)}$$

$$\begin{array}{ll} \min & EMD(X,Y) \\ \text{s.t.} & -\sum_{y=0}^{N-1} P_Y(y)\log_2(P_Y(y)) \geq R - H_X \\ & \sum_{x=0}^{M-1} P_X(x)P_{Y|X}(y|x) = P_Y(y), \forall y \\ & \sum_{y=0}^{N-1} P_{Y|X}(y|x) = 1, \forall x \\ & P_{Y|X}(y|x) \geq 0, \forall x, y \end{array} \tag{17}$$

where the $P_{Y|X}(y|x)$ is transition probability matrix, $\|d(x,y)\|^2$ structural similarity distortion measures. $P_X(x)$ is the constant parameters are the source distribution. Figuratively speaking, constraint $P_{Y|X}(y|x) \geq 0$ allows moving supplies from X to Y and not vice versa. Constraint $\sum_{x=0}^{M-1} P_X(x)P_{Y|X}(y|x) = P_Y(y)$ limits the clusters in Y to receive no more supplies than their weights, and the marginal distribution $P_Y(y)$ can be got by the fast algorithm in [8]. Constraint $\sum_{y=0}^{N-1} P_{Y|X}(y|x) = 1$ forces to move the maximum amount of supplies possible. We call this amount the total flow. Once the transportation problem is solved, and we have found the optimal flow $P_{Y|X}(y|x)$.

Property 2 (metric transitivity) : The ground distance $\|d(x,y)\|^2 = (\|x-y\|^2+c)/(\|x\|^2+\|y\|^2+c)$ is a metric and the total weights of the distributions X and Y are equal, then $EMD(X,Y)$ holds the property of metric.

Obviously, EMD(X,Y)=EMD(Y,X) and EMD(Y,X)$\geq$ 0, so we only need to prove that the triangle inequality holds [18]. Without loss of generality we consider the flow $X \Rightarrow Y \Rightarrow Z$. We assume supplies from x_i to y_j to z_k, then we have $\|d(x_i,y_j)\|^2 + \|d(y_j,z_k)\|^2 \geq \|d(x_i,z_k)\|^2$. Supposing $\hat{a}_{i,j}$ is an optimal matching to change X into Y, $\hat{b}_{j,k}$ is an optimal matching to change Y into Z and $h_{i,k}$ is an optimal matching to change X into Z. Consequently, the composite $h_{i,k}$ is derived from the $\hat{a}_{i,j}$ and $\hat{b}_{j,k}$ as the sum of interval intersections

$$h_{i,k} = \sum_{j=1}^{n}\left|\left[\sum_{i'=1}^{i-1}\hat{a}_{i',j}, \sum_{i'=1}^{i}\hat{a}_{i',j}\right] \cap \left[\sum_{k'=1}^{k-1}\hat{b}_{j,k'}, \sum_{k'=1}^{k}\hat{b}_{j,k'}\right]\right| \tag{18}$$

Because of the distributions X, Y and Z have equal weights. So we can proof as follow

$$\begin{aligned} EMD(X,Z) &\leq \sum_{i,k} h_{i,k}\|d(x_i,z_k)\|^2 \\ &\leq \sum_{i,j}\hat{a}_{i,j}\|d(x_i,y_j)\|^2 + \sum_{j,k}\hat{b}_{j,k}\|d(y_j,z_k)\|^2 \\ &= EMD(X,Y) + EMD(Y,Z) \end{aligned} \tag{19}$$

Therefore, the EMD is a true metric, which allows endowing optimal histogram modification with a metric structure.

The EMD(X,Y) can be modeled as a solution to a transportation problem, which is a special case of linear programming (LP) problems. One such efficient algorithm is the transportation simplex. As a general description of linear programming problem [2, 14].

$$\begin{aligned} \min\ & \mathbf{E}_{emd} = \mathbf{d}^T\mathbf{P} \\ \text{s.t.}\ & \mathbf{AP} = \mathbf{b} \\ & \mathbf{P} \geq 0 \end{aligned} \tag{20}$$

where $\mathbf{A} \in \mathbf{R}_{m\times n}$, $\mathbf{d} = [\|d(x,y)\|^2] \in \mathbf{R}_{n\times 1}$ and $\mathbf{P} = [P_{Y|X}(y|x)] \geq 0$. The optimization procedure of the simplex method is first illustrated with the assumption that the rows of $\mathbf{A}$ are linearly independent, which $\mathbf{A} = m$. So, the first m columns are assumed to be linearly independent and denoted as $\mathbf{A}_B$, and the other columns called as $\mathbf{A}_{NB}$. Then we have $\mathbf{A} = [\mathbf{A}_B, \mathbf{A}_{NB}]$. Further let $\mathbf{d} = \left[\mathbf{d}_B^T, \mathbf{d}_{NB}^T\right]^T$ and $\mathbf{P} = \left[\mathbf{P}_B^T, \mathbf{P}_{NB}^T\right]^T$, where $\mathbf{P}_B^T$ are basic variables and $\mathbf{P}_{nb}^T$ non-basic variables. Then the matrix description can be given by

$$\begin{pmatrix} \mathbf{1} & -\mathbf{d}_B^T & -\mathbf{d}_{NB}^T \\ \mathbf{0} & \mathbf{A}_B & \mathbf{A}_{NB} \end{pmatrix} \begin{pmatrix} \mathbf{E}_{emd} \\ \mathbf{P}_B \\ \mathbf{P}_{NB} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{b} \end{pmatrix} \tag{21}$$

Based on the above, the algebraic operations performed by the simplex method are expressed in matrix form by premultiplying both sides of the original set of equations by the appropriate matrix. Consequently, the desired matrix form of the set of equations after any iteration is

$$\begin{pmatrix} \mathbf{1} & \mathbf{0} & \mathbf{d}_B^T\mathbf{A}_B^{-1}\mathbf{A}_{NB} - \mathbf{d}_{NB}^T \\ \mathbf{0} & \mathbf{I}_m & \mathbf{A}_B^{-1}\mathbf{A}_{NB} \end{pmatrix} \begin{pmatrix} \mathbf{E}_{emd} \\ \mathbf{P}_B \\ \mathbf{P}_{NB} \end{pmatrix} = \begin{pmatrix} \mathbf{d}_B^T\mathbf{A}_B^{-1}\mathbf{b} \\ \mathbf{A}_B^{-1}\mathbf{b} \end{pmatrix} \tag{22}$$

As shown in [2, 14], the sufficient conditions which lead to the conclusion that the solution is global optimal are $\mathbf{A}_B^{-1}\mathbf{b} \geq \mathbf{0}$ and $\mathbf{d}_B^T\mathbf{A}_B^{-1}\mathbf{A}_{NB} - \mathbf{d}_{NB}^T \leq \mathbf{0}$. However, the above two requirements are not necessarily guaranteed. Firstly, the $\mathbf{A}_B^{-1}\mathbf{b} \geq \mathbf{0}$ is equivalent to state that $\mathbf{P} \geq \mathbf{0}$, which can be satisfied by introducing artificial variables. Secondly, $\mathbf{d}_B^T\mathbf{A}_B^{-1}\mathbf{A}_{NB} - \mathbf{d}_{NB}^T \leq \mathbf{0}$ needs an iterative procedure of switching basic and non-basic variables. Until it satisfies the requirement that $\mathbf{d}_B^T\mathbf{A}_B^{-1}\mathbf{A}_{NB} - \mathbf{d}_{NB}^T \leq \mathbf{0}$, a global optimal solution is reached. Otherwise, keep switching another pair of basic and non-basic variables, until an optimal result has been obtained. It is worth noting the above equation are based on the assumption that $\mathbf{A} = m$. However, it is not always $\mathbf{A} = m$ in some cases, and $\mathbf{A}_B$ does not exist inverse matrix $\mathbf{A}_B^{-1}$. Artificial variables described in [2, 14] is effectively method, which facilitates the selection of the m basic variables as well as solving the problem caused by redundant constraints. The modified form in equation can be given by

$$\begin{aligned} \min\ & \mathbf{E}_{emd} = \mathbf{d}^T\mathbf{P} + \gamma[1, 1, ..., 1]\mathbf{I}_m\mathbf{P}_A \\ \text{s.t.}\ & \mathbf{AP} + \mathbf{I}_m\mathbf{P}_A = \mathbf{b} \\ & \mathbf{P} \geq 0 \end{aligned} \tag{23}$$

where γ is an unspecified large positive number, if $\gamma > max(\mathbf{d}_T)$ will make solutions including nonzero artificial variables not the optimal solutions. $\mathbf{P}_A$ are referred to as artificial variables. Therefore, the constraint matrix changes to $[\mathbf{A}, \mathbf{I}_m]$ and $rank([\mathbf{A}, \mathbf{I}_m]) \geq m$, which can guarantee the inverse matrix $\mathbf{A}_B^{-1}$. Consequently, we can obtain the optimal solution $\mathbf{P} = [P_{XY}(xy)]$ to Eq.23 by using the procedures described in the previous subsection. Based on $P_{Y|X}(y|x)$, $P_X(x)$ and $P_Y(y)$, we can calculate the optimal transition probability matrix $P_X(x)P_{Y|X}(y|x) = P_{XY}(xy)$ and $P_{X|Y}(x|y) = P_{XY}(xy)/P_Y(y)$.

2.4 Recursive code construction for RDH

Recursive code construction (RCC) has first proposed by Kalker and Willems [8] and developed by Zhang et al. [31] for RDH. By RCC, we divide the host sequence into disjoint blocks and embed the message by modifying the histogram of each block. We first divide the host sequence x into g disjoint blocks, in which the first $g-1$ blocks have the same length K, and the last block has the length L_{last}, and thus $N = K(g-1) + L_{last}$. To finish the embedding, we have to set L_{last} to be larger than K. The ith cover block is denoted by x_i, and the corresponding stego block is denoted by y_i, $i = 1, \ldots, g$. We embed the message into each block by an embedding function $Emb()$, such that $(\mathbf{M}_{i+1}, y_i) = Emb(\mathbf{M}_i, x_i)$, with $i = 1, \ldots, g$ and $\mathbf{M}_1 = \mathbf{m}$ [5]. In other words, the embedding process in the ith block outputs the message to be embedded into the $(i+1)$th block. The $\mathbf{M}_{i+1}$ consists of the the rest message bits and the overhead information, $O(x_i)$, for restoring x_i. The message extraction and cover reconstruction are processed in a backward manner with an extraction function $Ext()$, such that $(\mathbf{M}_i, x_i)$=$Ext(\mathbf{M}_{i+1}, y_i)$, with $i = g, \ldots, 1$.

Now we consider a sender with a distortion constraint Δ. To maximize the embedding rate, we use EMD to estimate the optimal transition probability matrix $P_{Y|X}(y|x)$ of problem (14) according to Δ and the host distribution $P_X(x)$, and then we can calculate the transition probability matrix $P_{Y|X}(y|x)$. The embedding and extracting processes will be realized by the decompression and compression algorithms of an entropy coder (e.g., arithmetic coder) with $P_{Y|X}(y|x)$ and $P_{X|Y}(x|y)$ as parameters. We denote the compression and decompression algorithms by $Comp()$ and $Decomp()$ respectively. For simplicity, we assume Y is just a random variable satisfying the optimal marginal distribution $P_Y(y)$ that is determined by $P_{Y|X}(y|x)$ and $P_X(x)$. Therefore, the rate-distortion bound (14) can be rewritten as

$$\begin{aligned} \rho_{rev}(\Delta) &= maximize\{H(Y)\} - H(X) \\ &= H(Y) - H(X) \\ &= H(Y|X) - H(X|Y). \end{aligned} \tag{24}$$

On the other hand, in a K-length block of the code construction, we modify the host signal x to y according to the optimal transition probability $P_{Y|X}(y|x)$, so the average distortion d is given by $d = \sum_{x,y} P_X(x)P_{Y|X}(y|x)D(x, y)$. Note that $P_{Y|X}(y|x)$ is the solution of (14) under the condition $\sum_{x,y} P_X(x)P_{Y|X}(y|x)D(x, y) \le \Delta$, so we have $d \le \Delta$.

In a K-length block, the average number of embedded message bits is given by $\sum_{x=0}^{B-1} P_X(x)H(Y|X = x)$ and the average capacity cost for reconstructing this block is given by $\sum_{y=0}^{B-1} P_Y(y)H(X|Y = y)$, so the embedding rate R in one block is given by

$$\begin{aligned} R &= \sum_{x=0}^{B-1} P_X(x)H\left(Y|X = x\right) - \sum_{y=0}^{B-1} P_Y(y)H\left(X|Y = y\right) \\ &= H(Y|X) - H(X|Y). \end{aligned} \tag{25}$$

Thus, we get $R = \rho_{rev}(\Delta)$. Therefore, RCC can approach the ratedistortion bound (4) [26].

Data Embedding Process: The embedding is done by substituting signals of the cover with sequences obtained by decompressing the message bits in accordance with the the optimal transition probability matrix. In other words, for each bin x, $x \in \{0, \ldots, B-1\}$, we decompress a part of the message sequence according to the distribution $P_{Y|X}(y|x)$, and then substitute all host signals equal to x with the decompressed sequence. Thus, the histogram of the cover block is modified in a bin by bin manner. In each host block x_i, the embedding function $Emb()$ executes two tasks. One task is to embed some bits of the message and generate the stego-block y_i by decompressing the message sequence according to $P_{Y|X}(y|x)$. The other task is to produce the overhead information $O(x_i)$ for restoring the host block $\mathbf{x}_i$ by compressing it according to y_i and $P_{X|Y}(x|y)$. The overhead information will be embedded into the next block x_{i+1} as a part of $\mathbf{M}_{i+1}$ (see Fig. 2).

Data Extraction and Cover Restoration Processes: The data extraction and cover restoration are processed in a backward manner, such that $(\mathbf{M}_i, \mathbf{x}_i)$=$Ext(\mathbf{M}_{i+1}, \mathbf{y}_i)$ for $i = g-1, \ldots, 1$. From the $(i+1)$th stego block, we can extract the overhead $O(\mathbf{x}_i)$, by which we reconstruct the ith cover block $\mathbf{x}_i$. With the help of $\mathbf{x}_i$, we can extract the message from $\mathbf{y}_i$ by decompressing it according to the the optimal transition probability matrix $P_{Y|X}(y|x)$. In each stego block $\mathbf{y}_i$, the extraction function $Ext()$ also executes two tasks. One task is to decompress the overhead information extracted from $\mathbf{y}_{i+1}$ according to $P_{X|Y}(x|y)$ and restore the host block $\mathbf{x}_i$. The other task is to extract the message by compressing $\mathbf{y}_i$ according to $\mathbf{x}_i$ and $P_{Y|X}(y|x)$.

3 Application, experiment and analysis

3.1 Prediction and double-layered embedding method

In this paper, we employ double-layered embedding method [19]. All pixels are divided into two sets: the shadow pixel set (Dot) and the blank set (Five Star) (see Fig. 3). In the first round, the shadow set is used for embedding data and blank set for computing predictions, while in the second round, the blank set is used for embedding and shadow set for computing predictions. Since the two layers embedding processes are similar in nature, we only take the shadow layer for illustration.

Next, the prediction-error is computed by:

$$\mathbf{x^e} = P - \hat{P} \tag{26}$$

$$\hat{P} = \frac{p_{i-1,j} + p_{i,j+1} + p_{i+1,j} + p_{i,j-1}}{4} \tag{27}$$

Finally, the prediction-error sequence $\mathbf{x^e} = \{x_1^e, \cdots, x_N^e\}$ is derived.

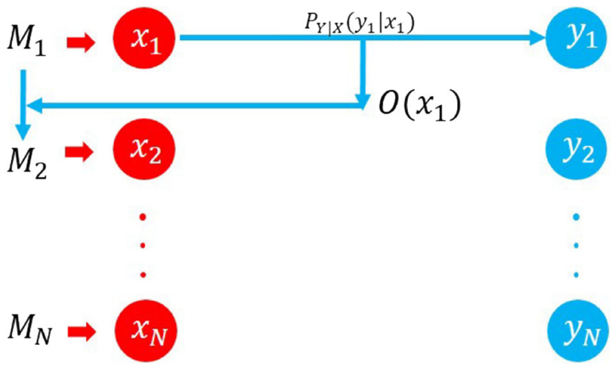


Fig. 2 Illustration of recursive code construction

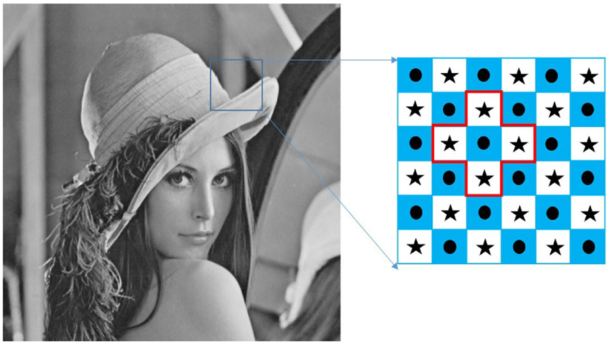


Fig. 3 The image divided into two sets: the Dot pixel set and the Five Star set

3.2 Experiment, analysis and comparison

In fact, the size of distortion metrics $\|d(x, y)\|^2$ is related to the size of prediction-error. We define $x^e_{min} = min\{x^e_1, \cdots, x^e_N\}$ and $x^e_{max} = max\{x^e_1, \cdots, x^e_N\}$, so the prediction-error range form x^e_{min} to x^e_{max}. In experiments, we truncate the prediction-error by this way: $x^e_{Th} = \{\mathbf{x^e} | \mathbf{x^e} \geq Th\}$, where $Th \geq |x^e_{min}|$. Therefore, the size of distortion metrics $\|d(x^e, y^e)\|^2$ is $(x^e_{max} - x^e_{Th}) \times (x^e_{max} - x^e_{Th})$.

The flow chart of embedding and extracting is in Fig. 4. The details Procedures of optimal structural similarity constraint as follows:

Algorithm 1 The Embedding Procedures of Optimal Structural Similarity Constraint (OSSC) for RDH.

Require:

1: The cover prediction-error $\mathbf{x^e}$ distribution, $P_{X^e}(x^e)$.
2: The embedding rate R.
3: The structural similarity distortion constraint $\|d(x^e, y^e)\|^2$.
4:

Ensure:

5:
6: Calculate the prediction-error sequence $\mathbf{x^e} = \{x^e_1, \cdots, x^e_i, \cdots, x^e_{k\times m}\}$ is obtained from the cover;
7: Use the EMD to calculate the optimal transition matrix $P_{Y^e|X^e}(y^e|x^e)$ and $P_{X^e|Y^e}(x^e|y^e)$ by the simplex method. Use the fast algorithm in [8] to get the optimal marginal distribution $P^e_Y(y^e)$.
8: Set block length K , the length of the last block L_{last} and then determine the number of blocks g.
9: Embed the message and the LSBs of the last block into the first $g - 1$ blocks with the RCC method.
10: Embed the overhead $O(x^e_{g-1})$ and parameters, including compressed frequencies of host,Δ, K and L_{last} into the last block by LSB replacement.
11: If $O(\mathbf{x}^e_{g-1})$ and the parameters can be completely embedded into the last block, output the stego sequence y^e; otherwise, set $L_{last} = L_{last} + K$ and $g = g - 1$, and redo Step 3–Step 5. **return** the marked prediction-error sequence $\mathbf{x^e} = \{x^{e\prime}_1, \cdots, x^{e\prime}_i, \cdots, x^{e\prime}_{k\times m}\}$.

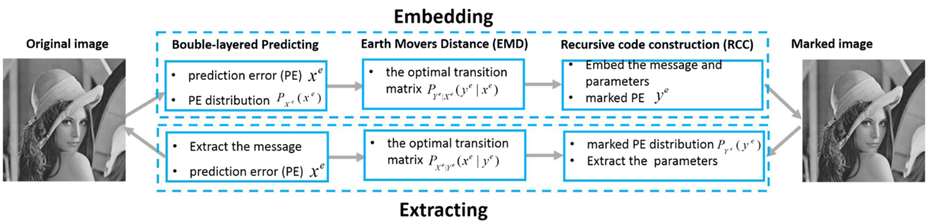


Fig. 4 The flow chart of OSSC

Algorithm 2 The Extracting Procedures of Optimal Structural Similarity Constraint (OSSC) for RDH.

Require:

The marked prediction-error distribution $P_{Y^e}(y^e)$.

The structural similarity distortion constraint $\|d(x^e, y^e)\|^2$.

Ensure:

Qbtain the marked prediction-error sequence $\mathbf{x^e} = \{x_1^{e\prime}, \cdots, x_i^{e\prime}, \cdots, x_{k\times m}^{e}{}'\}$.

Extract the length of the last block, L_{last}, from the LSBs of the 15 elements at the tail.

According to $P_X^e(x^e)$ and Δ, calculate the optimal transition matrices $P_{Y^e|X^e}(y^e|x^e)$ and $P_{X^e|Y^e}(x^e|y^e)$.

Implement the data extraction and host restoration procedure. Output the message **m**, the LSBs of the last block, and the first $g-1$ blocks of the host sequence $\mathbf{x^e}$. **return** Recover the original prediction-error sequence $\mathbf{x^e} = \{x_1^e, \cdots, x_i^e, \cdots, x_{k\times m}^e\}$.

In practice, one usually requires a single overall quality measure of the entire image. We use a mean SSIM (MSSIM) index to evaluate the overall image quality

$$MSSIM(X, Y) = \frac{1}{M}\sum_{i=1}^{M} SSIM(x_i, y_i) \tag{28}$$

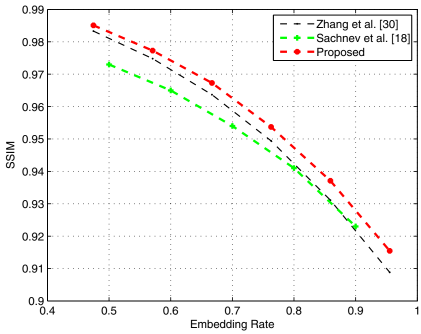
where X and Y are the reference and the distorted images, respectively. x_i and y_i are the image contents at the jth local window. M and is the number of local windows of the image.

We implemented these methods on the computer with Intel core i3 and 4GB RAM. The program developing environment is MATLAB R2011b based on Microsoft Windows 7 operating system. In the experiment, in order to simplify the complexity of OSSC, let $c = 200$. We implemented the proposed code construction with arithmetic coder as the entropy coder. In the experiment, we set the block length $K = 7000$ and the length of the last block $L_{last} = 4000$, and $Th = max\{400 - R \times 800, 10\}$ Test image is shown in Fig. 5. Besides, we select some images (Fig. 6) from the LIVE (Laboratory for Image and Video Engineering) [24] database to test our OSSC algorithm.

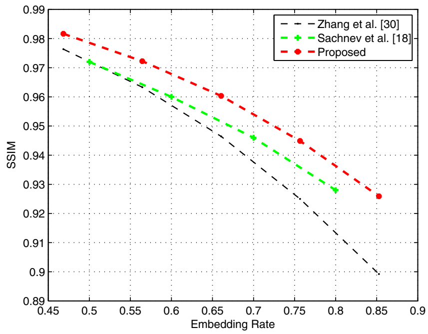
Observing from Fig. 7a, b, c and d, we compare our OSSC method with Zhang et al. [31]. Figure 7a, b, c and d illustrates that if embedding rate is larger, the effect brought by our algorithm is more obvious. Comparing with Zhang et al. [31] , our method have gained

Fig. 5 The test image Lena, barbara, cornfield and boat

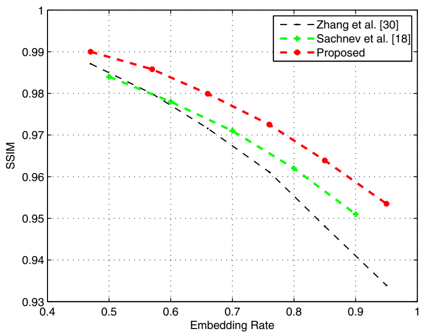
Fig. 6 The test image flowersonih, lighthouse, manfishing and carnivaldolls in LIVE database



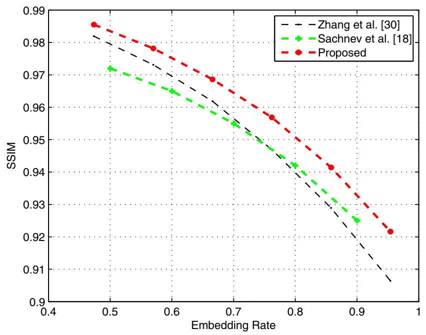
(a) lena



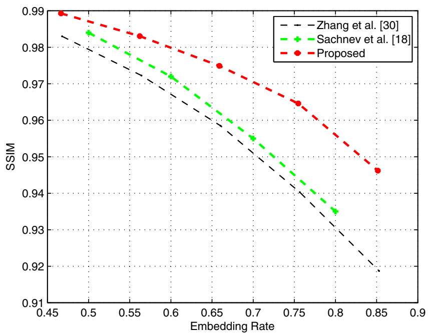
(b) barbara



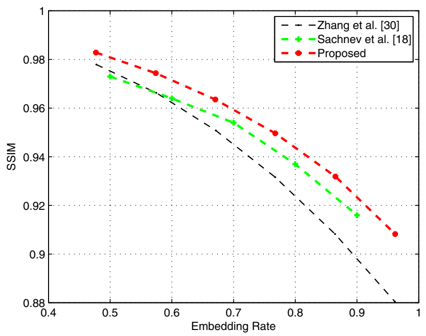
(c) cornfield



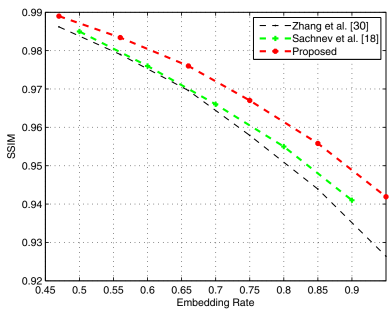
(d) boat



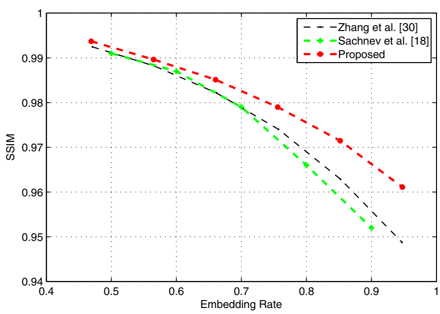
(e) flowersonih



(f) lighthouse



(g) manfishing



(h) carnivaldolls

Fig. 7 Embedding performance comparisons with Zhang et al. [31] and Sachnev et al. [19]

of MSSIM is much more higher, about 0.01 to 0.02 on average. Especially under large embedding rates. In Fig. 7e, f, g and h, we compare our OSSC method with Zhang et al. [31], which illustrates that if embedding rate is larger, the effect brought by our algorithm is more obvious. Comparing with Zhang et al. [31] , our method have gained of MSSIM is much more higher, about 0.02 to 0.03 on average. Especially under large embedding rates.

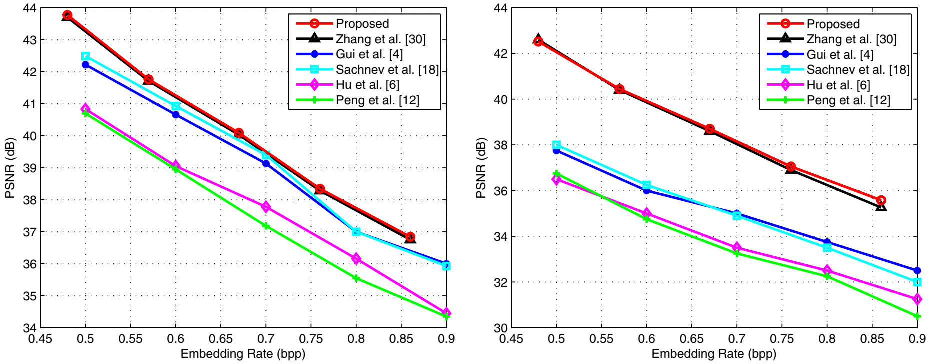
Observing from Table 1, we compare our method with Zhang et al. [31] method in different embedding rate. Comparing with the Zhang et al. [31] method, our OSSC method gains 0.77 %, 2.97 %, 2.14 %, 1.68 %, 1.24 %, 1.09 %, 3.00 %, 3.16 %, 1.67 %, 1.95 %, 1.34 %, 1.44 % in test image lena, barbara, corneld, boat, man, cablecar, owersonih, lighthouse, manshing, sailing, carnivaldolls, house, respectively. For our method, an average 1.89 % gains is earned compared with the Zhang et al. [31] method. Especially under high embedding rates. Both of this indicate that to some extent, the OSSC strategy for RDH is efficiency

Besides the MSSIM, our method can achieve good performance performance in the term of PSNR. Observing form Fig. 8, one can find our method is being compared with the other five recent works of Gui et al. [4], Sachnev et al. [19], Hu et al. [6], Peng et al. [13]. The comparison results are shown in Fig. 8a and b. In conclusion, compared with the state-of-the-art works [4, 6, 13, 19, 24], the superiority of the propose method is experimentally verified. It demonstrates the effectiveness of the proposed SSIM-based embedding strategy.

Reversible data hiding, as a fragile watermarking technique, is largely used for data integrity authentication, and data annotation. It requires the cover itself to be completely recovered after extracting the embedded message, which is very useful in fileds like medical imagery, military imagery and law forensics. In recent years, the hotspots of reversible data hiding has been directed to the capacity-distortion performance. Until now, almost reversible data hiding techniques competes each other by the capacity-distortion curve. Such as [3, 4, 7, 10–12, 16, 17, 19–21, 27, 31]. Reversible data hiding embeds messages into the smooth regions of an image because of good rate distortion compromise, while the statistical invisibility and security properties for reversible data hiding are some kinds of less concerned by most researchers. Actually the invisibility and security concerns are very good research directions for reversible data hiding, through which we can get better balance among the

Table 1 series Table 1: OSSC Embedding performance comparisons with Zhang et al. [31]

	Image	Embed-Rate	Zhang et al.	Proposed	Improvement	Increased Percentage
1	lena	0.96	0.9088	0.9158	0.0070	0.77 %
2	barbara	0.85	0.8992	0.9259	0.0267	2.97 %
3	cornfield	0.95	0.9339	0.9539	0.0200	2.14 %
4	boat	0.95	0.9064	0.9216	0.0152	1.68 %
5	man	0.90	0.9189	0.9303	0.0114	1.24 %
6	cablecar	0.95	0.9520	0.9624	0.0104	1.09%
7	flowersonih	0.85	0.9187	0.9462	0.0275	3.00 %
8	lighthouse	0.96	0.8804	0.9082	0.0278	3.16 %
9	manfishing	0.95	0.9264	0.9419	0.0155	1.67 %
10	sailing	0.96	0.8919	0.9093	0.0174	1.95 %
11	carnivaldolls	0.95	0.9487	0.9611	0.0124	1.34 %
12	house	0.96	0.9066	0.9197	0.0131	1.44 %



(a) Lena (b) Boat

Fig. 8 **a** and **b** is performance comparison between our method and six methods of Zhang et al. [31], Gui et al. [4], Sachnev et al. [19], Hu et al. [6], Peng et al. [13]

three properties and make reversible data hiding more applicable. We are highly encouraged to discuss deeply into these aspects in our later work.

4 Conclusion

In this paper, we utilize the characterize of SSIM to design RDH algorithm, and propose an optimal RDH algorithm under structural similarity constraint. Firstly, SSIM is often used as a black box in optimization tasks as merely an adhesive control unit outside the main optimization module. we deduce the metric of the structural similarity constraint, and further we prove it does't hold NCE property. Secondly, we construct the rate-distortion function under structural similarity distortion constraint, which can obtain the optimal transition probability matrix. Experiments show that our proposed OSSC method is effective.

Acknowledgments This work was supported in part by the Natural Science Foundation of China under Grants 61170234 and 60803155, and by the Strategic Priority Research Program of the Chinese Academy of Sciences under Grant XDA06030601.

References

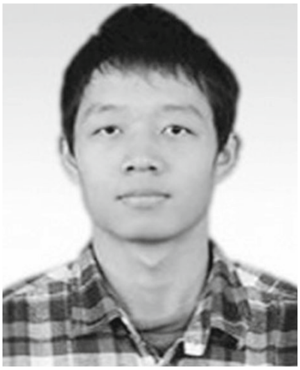
1. Brunet D, Vrscay ER, Wang Z (2012) On the mathematical properties of the structural similarity index. IEEE Trans Image Process 21(4):1488–1499
2. Cohen S (1999) Finding color and shape patterns in image. In: Ph.D. dissertation, Stanford University
3. Fridrich J, Goljan M (2002) Lossless data embedding for all image formats. Proc SPIE 4675:572–583
4. Gui X, Li X, Yang B (2014) A high capacity reversible data hiding scheme based on generalized prediction-error expansion and adaptive embedding. Signal Process 98(2):370–380
5. Haibin L, Kazunori O (2007) An Efficient Earth Movers Distance Algorithm for Robust Histogram Comparison. IEEE Trans Pattern Anal Mach Intell 29(5):840–853
6. Hu Y, Lee HK, Li J (2009) DE-based reversible data hiding with improved overflow location map. IEEE Trans Circuits Syst Video Technol 19(2):250–260
7. Hu X, Zhang W, Hu X, Yu N, Zhao X, Li F (2013) Fast estimation of optimal marked-signal distribution for reversible data hiding. IEEE Trans Inf Forensics Secur 8(5):779–788
8. Kalker T, Willems F (2002) Capacity bounds and code constructions for reversible data-hiding. In: Proc. of 14th International Conference on Digital Signal Processing, DSP2002, pp. 71–76

9. Laboratory for Image and Video Engineering: http://live.ece.utexas.edu/
10. Li X, Yang B, Zeng T (2011) Efficient reversible watermarking based on adaptive prediction-error expansion and pixel selection. IEEE Trans Image Process 20(12):3524–3533
11. Lin S-J, Chung W-H (2012) The Scalar Scheme for Reversible Information-Embedding in Gray-scale Signals: Capacity Evaluation and Code Constructions. IEEE Trans Inf Forensics Secur 7(4):1155–1167
12. Ni Z, Shi Y, Ansari N, Wei S (2006) Reversible data hiding. IEEE Trans Circuits Syst Video Technol 16(3):354–362
13. Peng F, Li X, Yang B (2012) Adaptive reversible data hiding scheme based on integer transform. Signal Process 92(1):54–62
14. Qi Z, Zhi Y, Hai T (2010) Differential earth movers distance with its applications to visual tracking. IEEE Trans Pattern Anal Mach Intell 32(2):274–287
15. Qin C, Zhang X (2015) Effective Reversible Data Hiding in Encrypted Image with Privacy Protection for Image Content. J Vis Commun Image Represent 31:154–164
16. Qin C, Chang C, Hsu T (2015) Reversible Data Hiding Scheme Based on Exploiting Modification Direction with Two Steganographic Images. Multimedia Tools and Appl 74(15):5861–5872
17. Qin C, Chang C, Huang Y, Liao L (2013) An Inpainting-Assisted Reversible Steganographic Scheme Using a Histogram Shifting Mechanism. IEEE Trans Circuits Syst Video Technol 23(7):1109–1118
18. Rubner Y (1999) Perceptual metrics for image database navigation. In: Ph.D. dissertation Stanford University
19. Sachnev V, Kim HJ, Nam J, Suresh S, Shi Y (2009) Reversible watermarking algorithm using sorting and prediction. IEEE Trans Circuits Syst Video Technol 19(7):989–999
20. Tian J (2003) Reversible data embedding using a difference expansion. IEEE Trans. Circuits Syst Video Technol 13(8):890–896
21. Thodi DM, Rodriguez JJ (2007) Expansion embedding techniques for reversible watermarking. IEEE Trans Image Process 16(3):721–730
22. Wang Z, Bovik AC (2009) Mean squared error: Love it or leave it? A new look at signal fidelity measures. IEEE Signal Process Mag 26(1)
23. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP (2004) Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 13(4)
24. Wang S, Rehman A, Wang Z, Ma S, Gao W (2013) SSIM-inspired divisive normalization for perceptual video coding. IEEE Trans Image Process 22(4):1418–1429
25. Xiannian F, Changhe Y, Malone BM (2014) Tightening bounds for bayesian network structure learning. AAAI
26. Xiannian F, Malone B, Yuan C (2014) Finding optimal Bayesian network structures with constraints learned from data. In: Proceedings of the 30th Annual Conference on Uncertainty in Artificial Intelligence (UAI-14)
27. Zhang X (2013) Reversible Data Hiding with Optimal Value Transfer. IEEE Trans Multimedia 15(2):316–325
28. Zhang W, Chen B, Yu N (2011) Capacity-Approaching Codes for Reversible Data Hiding. In: Proc. of 13th Information Hiding Conference, Prague
29. Zhang W, Chen B, Yu N (2011) Capacity-approaching codes for reversible data hiding. In: Proceeding 13th Information Hiding Conference, LNCS 6958, pp 255–269
30. Zhang W, Chen B, Yu N (2012) Improving Various Reversible Data Hiding Schemes via Optimal Codes for Binary Covers. IEEE Trans Image Process 21(6):2991–3003
31. Zhang W, Hu X, Li X, Yu N (2013) Recursive Histogram Modification: Establishing Equivalency Between Reversible Data Hiding and Lossless Data Compression. IEEE Trans Image Process 22(7):2775–2785

Jiajia Xu received the B.S. degree in 2009 from the Hefei University of Technology (HFUT), and the M.S. degree in 2012 from the University of Science and Technology of China (USTC), Hefei, China. he is now pursuing the Ph.D. degree in USTC. His research interests include cloud security, image and video processing, video compression and information hiding.

Weiming Zhang received the M.S. degree and Ph.D. degree in 2002 and 2005, respectively, from the Zhengzhou Information Science and Technology Institute, Zhengzhou, China. Currently, he is 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.

Ruiqi Jiang received the B.S. degree in 2010 from the Haerbin Institute of Technology (HIT), and the M.S. degree in 2013 from the New York University. He is now pursuing the Ph.D. degree in University of Science and Technology of China (USTC), Hefei, China. His research interests include cloud security and information hiding.



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

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