IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON RECENT ADVANCES IN VIDEO CODING AND SECURITY

Received January 31, 2019, accepted February 20, 2019, date of publication March 5, 2019, date of current version April 2, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2903079

Unconstraint Optimal Selection of Side Information for Histogram Shifting Based Reversible Data Hiding

JUNXIANG WANG[1], XIN CHEN[1], AND YUNQING SHI[2], (Fellow, IEEE)
[1]Jingdezhen Ceramic Institute, Jingdezhen 333403, China
[2]New Jersey Institute of Technology, Newark, NJ 07102, USA

Corresponding author: Junxiang Wang (wjx851113851113@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61762054, in part by the Major Program of Natural Science Foundation of Jiangxi Province under Grant 20161ACB21009, in part by the National Science Foundation for Distinguished Young Scholars of Jiangxi Province under Grant 20171BCB23072, in part by the Major Program of Science and Technology Program of Jiangxi Provincial Education Department under Grant GJJ1707619, and in part by the Foundation of China Scholarship Council under Grant 201708360076.

ABSTRACT Histogram shifting (HS) as a typical reversible data hiding (RDH) scheme is widely researched due to its high quality of stego-image. During HS process, the selected side information, i.e., peak and zero bins, usually greatly affects the performance of stego-image. Due to the massive solution space and burden in distortion computation, conventional HS-based schemes commonly utilize some empirical criterions associated with many artificial-designed constraints to determine side information, which could not lead to a global optimal performance for HS-based RDH. Later, our previous work proposed an adaptive information selection scheme for ''multiple embedding'' method by removing lots of constraints employed in conventional schemes. However, those constraints could not be completely avoided. Those chosen peak and zero bins in ''multiple embedding'' process are required to be different with each other, since they are commonly determined at one time as side information. In this paper, we employ ''multilevel embedding'' method and contrive to get rid of these above-mentioned constraints, which are called ''unnecessary constraints'' in this paper, so as to search essential global optimal side information around the entire solution space for HS-based RDH. Apparently, the process will dramatically increase the solution space and computation complexity. To effectively control the time cost, two novel approaches are proposed: 1) A pattern-based rapid performance evaluation method is designed to compute the rate and distortion; 2) Spirited by our previous work, a problem-oriented designed evolutionary algorithm, i.e., transfer learning-based genetic algorithm, is proposed to perform high-efficiency search around the entire huge solution space. For a given data payload, the proposed scheme could adaptively determine the optimal combination of peak and zero bins without any unnecessary constraint. The experimental results demonstrate the superiority of the proposed scheme compared with other state-of-the-art methods.

INDEX TERMS Reversible data hiding, histogram shifting (HS), unconstraint optimal selection, optimal peak and zero bins, genetic algorithm (GA), transfer learning.

I. INTRODUCTION

With the development of computer and network techniques, many multimedia are transmitted [1]–[2] and processed [3]–[5] via the internet. Meanwhile, to protect their copyright and authenticate the content integrity for those transitive data, some security techniques, such as reversible data hiding technique (RDH) [6]–[39], are widely researched in recently years. It could enable the decoder to not only extract the secret data as traditional schemes, but also perfectly reconstruct the original cover image without any distortion. For some specific scenarios, such as military, medical and legal applications, even slight distortion in images is not tolerated and thus RDH is perfectly employed in those cases.

In general, the existing reversible data hiding schemes can be mainly classified into several categories: i.e., lossless

The associate editor coordinating the review of this manuscript and approving it for publication was Zhaoqing Pan.

compression [6]–[7], difference expansion (DE) [8]–[16], histogram shifting (HS) [17]–[34] and others [35]–[39]. In addition, to ensure those data security in the cloud environment, some reversible information hiding algorithms are extended into encryption domain [40]–[41]. Among them, both DE and HS attract much attention due to their easy operation and high quality of stego-image. Initial DE based scheme [8] calculates the difference between two adjacent pixels and then doubles (expands) the value to vacate its LSB to hide 1-bit secret data. To further exploit image correlation, DE is extended on prediction errors, and called as prediction-error expansion (PEE) [9]–[15]. Later, several improved techniques, such as selection techniques [12], [16], location map reduction [14], are proposed on PEE. Recently, in [23] and [42], the ideal that DE/PEE scheme could be regarded as a special case of HS is proposed. Therefore, we concentrates on the HS based scheme in this paper.

Histogram shifting (HS) based scheme was initially presented by Ni *et al.* [17]. It selects a pair of peak bin (with highest frequency) and zero bin (with zero frequency) in histogram as side information and then shifts those bins between peak and zero bins by 1 towards zero bin to create vacant space nearby the peak bin. Finally, each pixel at peak bin is employed to embed 1-bit secret message for reversible data hiding. The process is simply illustrated by Fig.1(a) and the result is shown in Fig.1(b). It is also can be simply represented by Fig.2 by ignoring the vertical axis, where each value in the parenthesis means the frequency of one bin in current histogram.

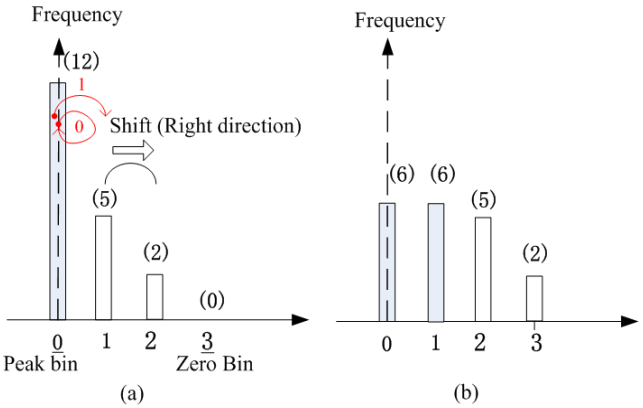


FIGURE 1. Illustration of histogram shifting (HS) based reversible data hiding process. (a) The process of HS. (b) Result of HS.

The flowchart of optimal rate allocation among multiple histograms

As shown in Fig.1, the rate (capacity) of HS based RDH is determined by the frequency of peak bin and the main distortion is caused by the shifted content between peak and zero bins. Generally speaking, the rate and distortion performance of HS based reversible data hiding heavily depends on ''the sharpness of histogram'' [18]–[31] and ''chosen side information'' [27], [30], i.e., peak and zero bins.

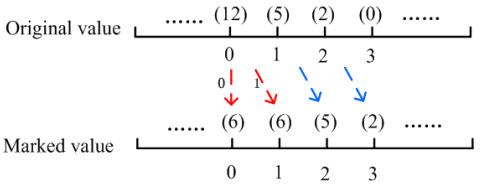


FIGURE 2. The simple representation of histogram shifting (HS) process of Figure 1.

So as to build sharp one histogram (or multiple histograms), some improvements are proposed, i.e., various accurate prediction and interpolation approaches [19]–[31] to build sharp histogram, sorting and selection scheme [19]–[30] to choose those pixels in the smooth regions for building sharp sub-histogram, and recent multiple histograms based schemes [32]–[35] to construct several tight/sharp sub-histograms according to their texture similarity.

In this paper, we focus on another issue. That is how to determine the optimal side information, i.e., combination of peak and zero bins, associated with the least distortion for the given payload. According to our understanding, the optimal selection of peak and zero bins could be modeled as an optimization problem as shown in (9), in which the fewer constraints are offered, the more flexible search is performed in a larger solution space. As a result, a better solution could be achieved in theory. Herein, based on search flexibility and the number of constraints in the optimization model, recent schemes could be simply summarized into three levels in this paper.

A. LEVEL 1: EMPIRICAL SEARCH OF SIDE INFORMATION

Ni *et al.* [17] employed the highest frequency bin as peak bin regardless of how much payload should be hidden. Latter, considering the Laplace distribution of prediction errors (PEs), Li *et al.* [31] and Sachnev *et al.* [19] generally utilized two fixed values at the center region of PE histogram, i.e. 0, -1, as peak bins for data hiding. In 2012, Xuan *et al.* [27] offered a flexible scheme, which found that combination of those bins with less frequency away from the center region of PE histogram as peak bins rather than those ones located at center region, i.e., 0, -1, might be a better choice in some cases. However, in [27], to control the computation complexity, peak bins are mandatory to be continuously selected, which means the case with continuous values, i.e., $\{-2, -3, -4\}$ rather than the combination of random determined values, i.e., $\{-2, -4\}$, is allowed to be peak bins, even the latter one might be better. Afterwards, Hwang [20] incorporated above-mentioned flexible side information selection method in Xuan *et al.* [27] and Sachnev*et al.*'s prediction [19] to offer an improved scheme. Based on the above analysis, it is clear that those empirical methods commonly add some additional constraints to achieve a rapid search, which is hardly to achieve global optimal side information.

B. LEVEL 2: "MULTIPLE EMBEDDING" WITH DIFFERENT (NON-REPEATED) PEAK BINS

To reduce the number of those constraints in empirical schemes (Level 1), our previous work [30] employed multiple embedding method with much more flexible selection approach to determine peak bins, which achieves a better performance. However, some inherent constriction for multiple embedding method could not be completely removed. It requires all the peak and zero bins should be chosen at one time and those values are different from each other, which is defined as ''non-repeated (also called single) peak bins'' case. The scheme could be illustrated in Fig.4(b) and Fig.5(a).

C. LEVEL 3: "MULTILEVEL EMBEDDING" WITHOUT THOSE (UNNECESSARY) CONSTRAINTS ON REPEATABILITY EMPLOYED IN LEVEL 2

To further remove the constraint on repeatability in multiple embedding method, multilevel embedding [34] is recommended in this paper, which allows to freely choose peak bins during the embedding process and thus some peak bins might be repeatedly utilized. As shown in Fig.5(b), multilevel embedding is illustrated, where peak bin 1 chosen in Layer 2 is actually generated from bin 0 in Layer 1, namely original bin 0 in Layer 1 is double (repeatedly) utilized in ''multilevel embedding''.

This paper focuses on the search of optimal side information in ''multilevel embedding'' process on Level 3. It is expected to obtain the essential global optimal solution. However, the process might lead to a drastically increased computation complexity due to two factors, i.e., the complex performance evaluation for multilevel embedding and greatly increased solution space due to less constraints. To deal with both issues (factors) and effectively control computation complexity, some techniques are designed and the contributions of this paper can be summarized as follows.

(1) Propose a more flexible side information selection approach with less constraints for ''multilevel embedding'' based RDH framework without those unnecessary constraints included in Level 1 and 2;

(2) Offer a pattern based rapid performance evaluation method for multilevel embedding process to reduce computation complexity;

(3) Design a novel problem-oriented evolutionary algorithm, i.e., Transfer Learning based Genetic Algorithm (GA), to achieve high-efficiency search of optimized side information within an entire huge solution space.

The rest of the paper is organized as follows. Some related work is briefly reviewed in Section II and the advantage of multilevel embedding process is discussed in Section III. Afterwards, two effective algorithms, i.e., 'rapid performance evaluation for multilevel embedding' and 'an evolutionary algorithm based effective search method', are designed to determine the optimal side information in Section IV and V, respectively. The embedding and extraction schemes are described in Section VI, which are followed by the experimental results and analysis in Section VII. Finally, the conclusion is offered in Section VIII.

II. RELATED WORKS

In the section, conventional HS based RDH process and their performance evaluation methods are simply reviewed, which will inspirit our design of rapid performance evaluation scheme for ''multilevel embedding'' in Section IV.

A. CONVENTIONAL HISTOGRAM SHIFTING (HS) BASED RDH

The first step of HS based RDH is to generate a sharp histogram, i.e., prediction error (PE) histogram. For each pixel x_i in cover image, it could be predicted by one precise prediction method with its neighborhood ones, i.e., precise rhombus prediction in [19], to obtain $\widehat{x_i}$ and then the prediction error e_i would be computed by

$$e_i = x_i - \widehat{x_i}. \tag{1}$$

Thus, all the e_i will be arranged in the raster scan order to generated an array as cover, denoted as $\mathbf{PE} = e_i | i \in [1, size]$, where $size$ is the number of e_i.

After **PE** generation, the HS embedding procedure is performed to obtain marked prediction error $\widetilde{e_i}$. First, for a histogram of **PE** (PEH), a peak bin with highest frequency and a zero bin with zero frequency in the histogram should be determined and denoted as (P_1, Z_1). HS then shifts the histogram bins between P_1 and Z_1 towards Z_1 direction by one unit to create a vacant position near P_1. Finally, each predictive error e_i is scanned to embed 1-bit message w when P_1 is encountered. When $P_1 < Z_1$, HS is performed towards 'right direction' as shown in Fig. 1 and computed by

$$\widetilde{e_i} = \begin{cases} e_i + 1, & e_i \in [P_1 + 1, Z_1 - 1] \\ e_i + w, & e_i = P_1 \\ e_i, & otherwise. \end{cases} \tag{2}$$

If $P_1 > Z_1$, HS embedding is similarly implemented towards the 'left direction'.

Finally, the marked pixel $\widetilde{x_i}$ is generated by $\widetilde{x_i} = \widehat{x_i} + \widetilde{e_i}$ to generate the stego-image.

During the process, notice two points, which improve the adaptability of HS based RDH and affect its performance.

(1) Peak bin P_1 is not obligatory to be highest frequency bin, but the appropriate one with enough frequency to hide given secret message.

(2) When relative large payload is required, rather than using just one pair of peak and zero bins (P_1, Z_1), some other strategies, i.e., 'multiple embedding' on Level 2 or 'multilevel embedding' on Level 3, might be employed, which could involve several pairs of peak and zero bins from the histogram and then perform multi-layer embedding.

As mentioned in Section I, the difference between both ''multiple embedding'' and ''multilevel embedding'' methods is the limitation whether peak bins could be repeatedly utilized in the embedding process. An example in Fig.5 is offered to illustrated their difference.

B. PERFORMANCE EVALUATION FOR CONVENTIONAL HS BASED EMBEDDING PROCESS

According to the process in above subsection, the performance for conventional HS based RDH, i.e., embedding capacity (also called 'rate' in this paper) and distortion, can be readily estimated as follows.

For single layer embedding with (P_1, Z_1) as peak and zero bins, the embedding capacity EC is as follows.

$$EC = h(P_1). \tag{3}$$

where $h(i)$ denotes the frequency of occurrence for value i in histogram.

Assume the '0' and '1' in the secret message of length C are equally distributed, the distortion D between original and stego pixels can be computed by the shifting value of each bin in the histogram and computed as

$$\begin{aligned} D &= D_{non-peak} + D_{signle-peak} \\ &= \sum_{i\in[P_1+1,Z_1-1]} h(i) \times (\Delta i)^2 + \sum_{'i=P_1' \text{and}' w=1'} (\Delta i)^2 \\ &= \sum_{i\in[P_1+1,Z_1-1]} h(i) + \frac{1}{2}C, \ (\Delta i = 1) \end{aligned} \tag{4}$$

where $\Delta(i) = 1$ means the shift value of each shifted bin is 1. The distortion is calculated in two parts: (a) The former term in (4) is the shifting distortion for non-peak bins, denoted as $D_{non-peak}$, caused by those bins' shifting as an entirety in the range $[P_1 + 1; Z_1 - 1]$; (b) The other one is the embedding distortion for peak bin P_1, denoted as $D_{single-peak}$, where means peak bins P_1 is only once (single) utilized and half of P_1 is shifted to its adjacent vacant value by 1.

Similarly, for ''multiple embedding'' on Level 2 involved with m pairs of different peak and zero bins, denoted as $\{(P_k, Z_k|1 \le k \le m)\}$, the total embedding capacity is calculated by

$$EC = \sum_{k=1}^{m} h(P_k). \tag{5}$$

The complete distortion D for m-layer multiple embedding can be rapidly evaluated by

$$\begin{aligned} D &= D_{non-peak} + D_{single-peak} \\ &= \sum_{'i\in[-255,255]' \text{and}' i\neq\{P_k|k\in[1,m]\}'} h(i) \times (\Delta bin_i)^2 + \sum_{k=1}^{m} D_{P_k} \end{aligned} \tag{6}$$

where Δbin_i means i-th bin's accumulated shifts value during m-layer multiple embedding. The distortion also is divided by two parts: (a) The former is the shifting distortion for those non-peak bins, which are shifted as an entirety with the same accumulated shifting value Δbin_i; (2) D_{P_k} represents the distortion caused by single utilized (non-repeated utilized) peak bin P_k, half of which will be changed to be adjacent value towards Z_k to hide $w = 1$ with shift being $(\Delta bin_{P_k} + 1)$ or $(\Delta bin_{P_k} - 1)$, which depends on the relative location between P_k and Z_k. Thus its distortion could be separately calculated as follows.

$$D_{P_k} = \begin{cases} \frac{1}{2} \times h(P_k) \times \left(\Delta bin_{P_k}\right)^2 + \frac{1}{2} \times h(P_k) \times \left(\Delta bin_{P_k} + 1\right)^2, \\ P_k < Z_k \\ \frac{1}{2} \times h(P_k) \times \left(\Delta bin_{P_k}\right)^2 + \frac{1}{2} \times h(P_k) \times \left(\Delta bin_{P_k} - 1\right)^2, \\ P_k > Z_k \end{cases} \tag{7}$$

During above performance estimation process, the accumulated shifts value Δbin_i for i-th bin in histogram is significant and could be calculated by

$$\Delta bin_i = \sum_{k=1}^{m} \Delta_k(i). \tag{8}$$

where $\Delta_k(i)$ means the shift value for the k-th layer embedding associated with P_k and Z_k as side information. It could be computed as

$$\Delta_k(i) = \begin{cases} 1, & i \in [P_k + 1, Z_k - 1] \\ 0, & otherwise \end{cases}$$

for the 'right direction' shifting when $P_k < Z_k$, or

$$\Delta_k(i) = \begin{cases} -1, & i \in [Z_k + 1, P_k - 1] \\ 0, & otherwise \end{cases}$$

for 'left direction' shifting when $P_k > Z_k$.

In addition, Δbin_{P_k} in (7) could be considered as a special case of $\Delta bin_{i=P_k}$ in (6), which menas the accumulated shifts value for peak bin P_k during the rest $(m-1)$-layer embedding where P_k is not regarded as the peak bin. The details could refer to our previous work [30].

Finally, it is noted that, since one peak bin is only single utilized in ''multiple embedding'', its half frequency will be changed to be the adjacent value as shown in Fig.1, and thus the embedding distortion will be readily computed by (7). On the contrary, if one peak bin is repeatedly employed in ''multilevel embedding'', the distortion calculation is much more complex and will be discussed in Section IV.

III. THE ADVANTAGE OF MULTILEVEL EMBEDDING WITHOUT ANY UNNECESSARY CONSTRAINTS

Compared with ''multiple embedding'' method, the superiority of ''multilevel embedding'' process employed in this paper is briefly discussed in terms of theory and practice.

A. THE ADVANTAGE OF THE MULTILEVEL EMBEDDING IN THEORY

Based on the given payload, our aim is to determine the optimal peak and zero bins as side information associated with minimum distortion. Based on the purpose, a general optimization model could be developed as follows.

$$\begin{cases} \underset{\{\mathbf{P},\mathbf{Z}\}}{Min}\ D(\mathbf{P},\mathbf{Z}) \\ s.t.\ h(\mathbf{P}) = \sum_{i=1}^{m} h(P_k) \ge C \\ \mathbf{P} \in \mathbf{Peak\_Set}, \quad \mathbf{Z} \in \mathbf{Zeak\_Set} \end{cases} \tag{9}$$

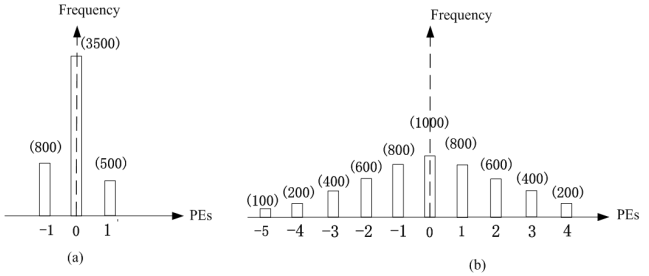


FIGURE 3. The artificial example: (a) Histogram corresponding to the smooth region. (b) Histogram corresponding to the texture region.

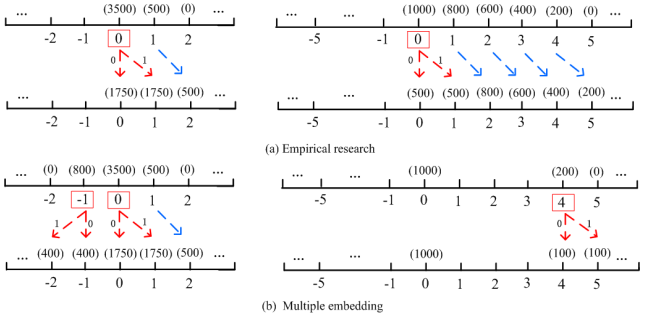


FIGURE 4. The selection of peak and zero bins by using different methods.

where D means the distortion. $\mathbf{P}$ and $\mathbf{Z}$ denotes the chosen optimal peak and zero bins involved in HS based RDH, and denoted as $\mathbf{P} = \{P_k | 1 \le k \le m\}$ and $\mathbf{Z} = \{Z_k | 1 \le k \le m\}$ respectively for m-layer embedding. C is the length of given secret message and constraint condition $\sum_{i=1}^{m} h(P_k) \ge C$ in (9) means given secret message could be completely embedded. All elements in $\mathbf{P}$ and $\mathbf{Z}$ are chosen from the candidate peak bins' set with all the non-zero frequency bins in the histogram, denoted as **Peak_Set**, and candidate zero bins' set including the rest zero frequency bins, denoted as **Zero_Set**, respectively. Clearly, D depends on the chosen $\mathbf{P}$ and $\mathbf{Z}$, and is expressed as $D(\mathbf{P},\mathbf{Z})$.

Evidently, both constraints in the optimization model (9) are inevitable for general RDH embedding. In addition, since other constraints mentioned in Level 1 and 2, which are defined as 'unnecessary constraints' in this paper, are not included in (9), it could achieve a most flexible search around the entire solution space and thus obtain the essentially global optimal side information in theory.

B. AN EXAMPLE TO DEMONSTRATE THE SUPERIORITY OF MULTILEVEL EMBEDDING

In this subsection, an artificial example is designed to illustrate some typical embedding methods on different levels (From Level 1 to Level 3 as mentioned in Section I), and then compare their performances.

As shown in Fig.3, assume the pixels in one cover image could be divided into both smooth and texture regions, respectively, where the smooth region corresponds to a sharp histogram with accurate prediction errors (PEs), denoted as H_a. While the texture region is associated with a fat histogram with lager PEs, denoted as H_b.

1) CASE 1: WHEN THE GIVEN PAYLOAD IS 4500 BITS

Level 1: As illustrated in Fig.4(a), the empirical search might fixedly choose two highest peak bins, e,g., bin 0 in histogram H_a and bin 0 in histogram H_b, respectively, to hide given payload. Its rate (embedding capacity) is 4500 bits according to (5), i.e., $EC = \sum_{k=1}^{2} h(P_k) = 3500 + 1000 = 4500$. Then, based on (6), (7) and Fig.4(a), its distortion is readily calculated by $D = [\frac{1}{2} \times 3500 \times (1-0)^2 + 500 \times (2-1)^2] + [\frac{1}{2} \times 1000 + 800 + 600 + 400 + 200] = 4750$, where the contents in each quare bracket mean the distortion caused by one histogram, i.e., H_a or H_b, and the underline is utilized to emphasize the distortion for peak bins.

Level 2: Similarly, as illustrated in Fig.4(b), for the multiple embedding, the bins $\{-1, 0\}$ in H_a, and bin 4 in H_b might be synchronously determined to hide secret data. Then, the rate is $EC = \sum_{k=1}^{3} h(P_k) = 800 + 3500 + 200 = 4500$ bits. The distortion is $D = [\frac{1}{2} \times 800 + \frac{1}{2} \times 3500 + 500] + [\frac{1}{2} \times 200] = 2750$.

It is observed that, since the flexible side information is selected without those unnecessary constraints, the distortion D for "multiple embedding" on Level 2 is less than that for empirical one on Level 1 at the same payload 4500 bits.

In addition, the solution space of "multilevel embedding" on Level 3 contains that of "multiple embedding" according to its fewer constraints. Thus, "multilevel embedding" could also reach the performance of "multiple embedding".

C. CASE 2: WHEN THE PAYLOAD IS INCREASED TO BE 8000 BITS

Level 2: As illustrated in Fig.5(a), for the multiple embedding, the bins $\{-1, 0, 1\}$ in H_a together with bins $\{-4, -1, 0, 2, 3, 4\}$ in H_b might be synchronously selected to hide secret data. Then, $EC = \sum_{k=1}^{9} h(P_k) = 8000$ bits and the distortion is $D = [\frac{1}{2} \times 800 + \frac{1}{2} \times 3500 + \frac{1}{2} \times 500 + \frac{1}{2} \times 500 \times (3-1)^2] + [\frac{1}{2} \times 200 \times (2)^2 + \frac{1}{2} \times 200 + \frac{1}{2} \times 800 + \frac{1}{2} \times 1000 + \frac{1}{2} \times 600 + \frac{1}{2} \times 600 \times 2^2 + \frac{1}{2} \times 400 \times 2^2 + \frac{1}{2} \times 400 \times 3^2 + \frac{1}{2} \times 200 \times 3^2 + \frac{1}{2} \times 200 \times 4^2 + 100 \times (2)^2 + 400 + 600 + 800] = 13600$

Level 3: As shown in Fig.5(b), for the multilevel embedding, the bins $\{0, 1, -1\}$ in H_a, and bins $\{-4, -1, 1, 4\}$ in H_b might be chosen to hide given payload.

It is observed that some peak bins might be repeatedly utilized in the multilevel embedding process since those peak bins in "multilevel embedding" are dynamically selected during each layer embedding process. As shown in the left

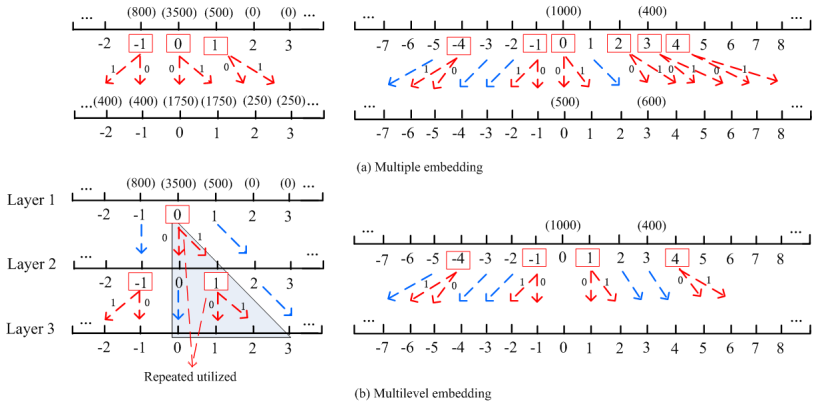


FIGURE 5. The selection of peak and zero bins.

picture in Fig.5(b), bin 1 in Layer 2 in histogram H_a originated from bin 0 in Layer 1 is also employed, namely original bin 0 in Layer 1 are double (repeatedly) utilized. In general, to evaluate the rate and distortion for multilevel embedding, the time-consuming embedding process should be tracked and the rate is $EC = \sum_{k=1}^{3} h(P_k) = 3500 + \frac{1}{2} \times 3500 + 800 + 200 + 800 + 800 + 200 = 8050$. Its corresponding distortion is $D = [\frac{1}{2} \times 800 + \frac{1}{2} \times \frac{1}{2} \times 3500 + \frac{1}{2} \times \frac{1}{2} \times 3500 \times 2^2 + 500 \times 2^2] + [\frac{1}{2} \times 200 \times (2)^2 + \frac{1}{2} \times 200 + \frac{1}{2} \times 800 + \frac{1}{2} \times 800 + \frac{1}{2} \times 200 + \frac{1}{2} \times 200 \times (2)^2 + 100 \times (2)^2 + 400 + 600 + 600 + 400] = 10975$

It is observed that "multilevel embedding" on Level 3 leads to less distortion compared with "multiple embedding" on Level 2 due to the flexible selection of some better side information.

Based on both examples, the superiority of "multilevel embedding" is demonstrated.

IV. RAPID PERFORMANCE EVALUATION OF MULTILEVEL EMBEDDING

To fast solve the optimization model (9) and thus determine the optimal side information for "multilevel embedding", a pattern based rapid performance evaluation method is proposed in this section to evaluate the performance of each solution in (9).

Inspirited from that method for "multiple embedding" as mentioned in Section II-B and comprehensive consideration of "multilevel embedding" process, the distortion D for the "multilevel embedding" can be firstly divided into three parts:

(1) Distortion caused by those non-peak bins, which are shifted as entirety (denoted as $D_{non-peak}$);

(2) Distortion caused by single used peak bins (denoted as $D_{single-peak}$);

(3) Distortion caused by repeatedly used peak bins (denoted as $D_{repeated-peak}$).

Thus the distortion D could be calculated by

$$D = \sum_{'i \in U[-255,255]' \text{and}' i \in \underline{non-peakbin}'} D_{non-peak} + \sum_{'i \in U[-255,255]' \text{and}' i \in \underline{single-peakbin}'} D_{single-peak} + \sum_{'i \in U[-255,255]' \text{and}' i \in \underline{repeated-peakbin}'} D_{repeated-peak} \tag{10}$$

Then, each part of distortions in (10) could be individually calculated as follows. Clearly, the front two items in (10), namely $D_{non-peak}$ and $D_{single-peak}$, could be easily computed by referring to multiple embedding process as mentioned in Section II-B.

A. CALCULATE $D_{NON-PEAK}$

If i-th bin in the histogram belongs to non-peak bin shifted as an entirety with accumulated shift value Δbin_i, $D_{non-peak}$ is computed by

$$D_{non-peak} = h(i) \times (\Delta bin_i)^2 \tag{11}$$

As shown in the left picture of Fig.5(b), denoted as Fig.5(b)_left, original bin '1' in Layer 1 belongs to this case. It is computed by $D_{i=1} = h(i) \times (\Delta bin_i)^2 = 500 \times (3-1)^2$.

B. CALCULATE $D_{SINGLE-PEAK}$

If one bin is a single used peak bin, denoted as P_k, with half of them changed to the adjacent value, $D_{single-peak}$ could be calculated by using (7).

As shown in Fig.5(b)_left, bin '-1' in Layer 1 belongs to the case, which is single utilized in the Layer 2 and half of $P_k = -1$ will be changed to be adjacent value -2. Thus, its distortion is calculated by $D_{i=-1} = \frac{1}{2} \times 800 \times (-1)^2 + \frac{1}{2} \times 800 \times (0)^2$.

C. CALCULATE $D_{REPEATED-PEAK}$

If one bin belongs to a repeatedly utilized peak bin (a particular case for multilevel embedding), a pattern based rapid

performance evaluation method is proposed in this paper to calculate the $D_{repeated-peak}$.

The main ideal is that all the possible patterns associated with their performances are pre-prepared in advanced. When multilevel embedding is encountered, one corresponding pre-prepared pattern could be rapidly searched to obtain its performance without repeatedly tracking the time-consuming multilevel embedding process.

Without loss of generality, assume that one repeatedly utilized peak bin is denoted as P_k associated with the number of reuses (namely repeated times) RT_P_k. Two patterns based on $RT\_P_k = 2$ are illustrated in Fig.6 and their usage are described as follows. Based on the Pattern 1 as shown in Fig.6(a) for 2-layer multilevel embedding process, the rate is the sum of the frequencies of twice used P_k and represented by $EC = h(P_k) + \frac{1}{2} \times h(P_k) = \frac{3}{2} \times h(P_k)$. The distortion for P_k could be easily calculated as $D_{repeated-peak} = \frac{1}{2} \times \frac{1}{2} \times h(P_k) \times (1)^2 + \frac{1}{2} \times h(P_k) \times (2)^2$. The performance results could be readily extended to m-layer ''multilevel embedding'' process and represented by

$$\begin{cases} EC = \dfrac{3}{2} \times h(P_k) \\ D_{repeated-peak} = \dfrac{1}{2} \times \dfrac{1}{2} \times h(P_k) \times (\Delta bin_{P_k} + 1)^2 \\ + \dfrac{1}{2} \times h(P_k) \times (\Delta bin_{P_k} + 2)^2 \end{cases} \quad (12)$$

where Δbin_{P_k} means the accumulated shift values for bin P_k shifted as entirety during the rest $(m-2)$-layers embedding, in which P_k is not considered as a peak bin.

Similarly, the performance, i.e., rate EC and distortion $D_{repeated-peak}$, for Pattern 2 in Fig.6(b) could be also are calculated by

$$\begin{cases} EC = \dfrac{3}{2} \times h(P_k) \\ D_{repeated-peak} = \dfrac{1}{2} \times \dfrac{1}{2} \times h(P_k) \times (\Delta bin_{P_k} + 1)^2 \\ + \dfrac{1}{2} \times \dfrac{1}{2} \times h(P_k) \times (\Delta bin_{P_k} + 2)^2 \end{cases} \quad (13)$$

In this way, all those performance results, i.e., EC and $D_{repeated-peak}$, associated with the corresponding pattern numbers, could be similarly pre-prepared and listed in the table for looking up.

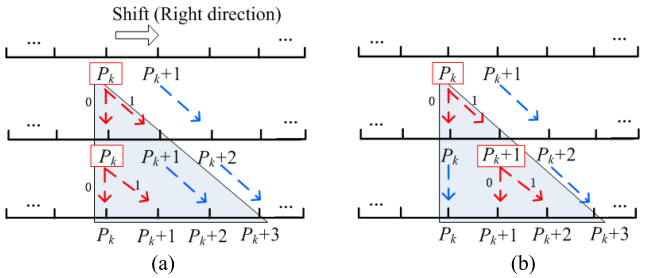


FIGURE 6. Two patterns based on $RT\_P_k = 2$.

Similarly, those patterns based on $RT\_P_k = 3$ associated with shifting towards ''right direction'' are listed in Appendix A for reference.

Finally, to calculate above three kinds of distortions in (10), the accelerate shifting value Δbin_i for i-th bin in histogram during m-layer multilevel embedding should be calculated, which might be present in the form of Δbin_{P_k} as a special case in (12) and (13) when $i = P_k$. In this paper, Δbin_i for ''multilevel embedding'' is similarly calculated as the case of ''multiple embedding'' process by using (8). The only difference is to supplement (8) by considering the special case of calculating $\Delta_k(i)$ for repeatedly utilized peak bins in multilevel embedding process as follows. When $P_k < Z_k$ for 'right direction' shifting,

$$\Delta_k(i) = \begin{cases} RT\_P_k, & i \in [P_k+1, Z_k-1] \\ 0, & otherwise \end{cases} \quad (14)$$

Otherwise, when $P_k > Z_k$ for 'left direction' shifting,

$$\Delta_k(i) = \begin{cases} -RT\_P_k, & i \in [Z_k+1, P_k-1] \\ 0, & otherwise \end{cases} \quad (15)$$

As illustrated in Fig.6 (a),

$$\Delta_k(i) = \begin{cases} 2, & i = P_k + 1 \\ 0, & otherwise. \end{cases}$$

According to above analysis, assume given side information for m-layer multilevel embedding process has s pairs of different peak and zero bins and could be presented as $P_k, k \in [1, s]$ with their corresponding number of reuse RT_P_k. Thus, $\sum_{k=1}^{s} RT\_P_k = m$ and $RT\_P_k \geq 1$. Then, the rapid performance evaluation could be performed as follows.

(1) Based on given side information, calculate the accelerate shifting value Δbin_i for each bin by using (8), (14), (15);

(2) Calculate distortions $D_{non-peak}$ and $D_{single-peak}$ by using (11) and (7), respectively;

(3) Determine the pattern number and then calculate $D_{repeated-peak}$ by looking up the corresponding performance in pre-prepared table, such as (12) or (13) for $RT\_P_k = 2$, or some performances listed in Appendix A for $RT\_P_k = 3$;

(4) Add those three parts of distortions to evaluate distortion D by (10);

(5) Add all the those frequencies of chosen peak bins to acquire rate EC.

By now, the performance evaluation is finished. It is clear that, without tracking the complicated multilevel embedding process, our scheme could save much time.

In addition, some issues are simply discussed.

(1) It is observed that Pattern 2 is better Pattern 1 for $RT\_P_k = 2$ due to the less distortion $D_{repeated-peak}$ based on the same rate EC, which will be employed for search of optimal side information in Section V.

(2) The number of pattern is limited, which is basis of pre-preparing all the possible patterns. Due to the requirement of imperceptible distortion of stego-image, the number of

reuses of peak bins RT_P_k is limited. According to our extended experimental results, RT_P_k is generally not more than 3. Based on the assumption that $RT\_P_k = 3$ and all the bins are shifted towards 'right direction', the number of different patterns is $1 \times 2 \times 3 = 6$ as shown in Appendix A. If two possible shifting directions for each bin are considered, the whole number of different patterns is $1 \times 2 \times 3 \times 2^{(3)} = 48$, among which some patterns are with the same performance and could be ignored.

V. AN EVOLUTIONARY ALGORITHM BASED EFFECTIVE SEARCH METHOD

As shown in optimization model (9), the solution space is huge and it is impossible to go through entire solution space to search the optimal solution by using empirical optimization methods, such as brute force search. Inspirited by our previous work [30], a heuristic evolutionary algorithm, i.e., Genetic Algorithm (GA), is employed to automatically determine those tunable parameters due to its rapid and stable convergence property. In addition, some special problem-oriented designs are proposed to further ensure a stable and high-efficiency search, i.e., (1) Transfer learning based initial seeds design; (2) Empirical evolution criterion design.

As shown in Fig.7, our genetic algorithm follows the conventional framework [43], which includes some problem oriented designs. Those seeds in initial population are randomly generated associated with some specifically designed empirical chromosomes with good genes to ensure better convergence for GA implementation. Let g, $G_{\max}$ and N denote the current generation number, the maximum generation number and the amount of individuals (seeds) in the initial population, respectively. The overall structure of the proposed genetic algorithm includes:

1) Initialization;

2) Empirical chromosomes (individuals) addition;

3) Evaluate all the individuals in the current population;

4) Divide those individuals into two parts, i.e., feasible part and infeasible one, according to the given payload constraint in (9);

5) Evolution. for each part, two ways are synchronously offered to achieve an expected evolution:

(a) Way 1: Conventional way including selection, crossover and mutation operations to generate the next generation, which could achieve a rapid search speed;

(b) Way 2: Specially designed empirical modification in order to achieve small search step and controllable search direction.

In addition, to ensure a stable and effective search results in such an entire huge solution space, transfer learning technique is optionally employed, which introduces some high-performance individuals, such as good individual from ''multiple embedding'' on Level 2 [30], into the initial population to guide evolution process. Those special design are described as follows.

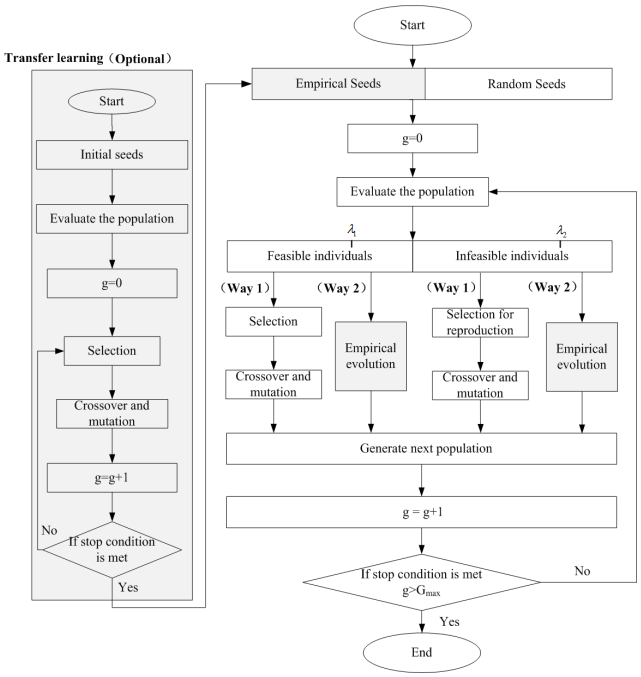


FIGURE 7. Flowchart of the proposed GA scheme.

A. CHROMOSOME ENCODING

To solve the optimization problem described in (9) using GA, all the tunable parameters, i.e., chosen peak and zero bins, the corresponding number of reuse RT_P_k, and pattern number are represented with chromosome encoding.

As shown in Fig.8, all the candidate peak bins from **Peak_Set** with non-zero frequency in the histogram are sorted in the ascending order and assume the number of elements in **Peak_Set** is denoted as q. Then each candidate peak bin, i.e., $P_k, k \in [1, q]$, is presented by the combination of four items, namely its value P_k, the number of reuse RT_P_k, corresponding zero bins Z_k from **Zero_Set** and the pattern number, denoted as PN_P_k, to denote how to repeatedly use P_k as mentioned in Section IV.

Apparently, if $RT\_P_k = 0$, P_k is not employed in the multilevel embedding process. When $RT\_P_k = 1$, P_k is single utilized similar as ''multiple embedding''. Then just one zero bin Z_k chosen from **Zero_Set** is employed and following pattern number could be ignored. When $RT\_P_k \geq 2$, the corresponding RT_P_k zero bins from **Zero_Set** associated with pattern number should be recorded.

B. TRANSFER LEARNING BASED INITIAL CHROMOSOMES ADDITION

Due to the entire huge solution space, completely random construction of initial population for GA might result in an unstable results, even the non-convergence one when the available solution space is tiny as mentioned in our previous work [30]. Therefore, it is desired to add some high-performance empirical seeds (individuals) into the ini-

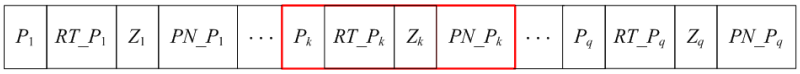


FIGURE 8. The encoder of each individual.

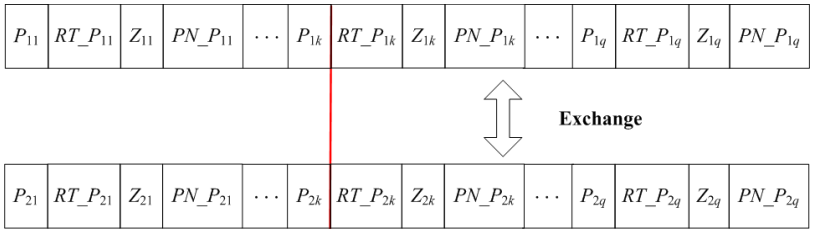


FIGURE 9. Crossover operation for two individuals.

tial pool of population to guide effective evolution and ensure a stable solution.

To achieve the purpose, transfer learning technique could be optionally employed into our GA based framework as shown in Fig.7, which will introduce an optimal solution for ''multiple embedding'' on Level 2 generated by our previous method [30].

In addition, as mentioned in [30], since the optimal solution search on Level 2 is rapid, the additional time cost for transfer learning technique is limited. The effectiveness of transfer learning technique will be discussed in Section VII.

C. EMPIRICAL EVOLUTION

Based on above analysis, completely random evolution might result in some uncontrollable convergence direction and unstable results. Based on this consideration, two ways, i.e., conventional method and a specially designed empirical modification, are simultaneously used to offer a controllable evolution scheme, where conventional evolution ensures convergence speed and empirical modification guides a controllable evolution direction with small search step size. The details are mentioned as follows.

1) CONVENTIONAL EVOLUTION

As shown in Fig.7, conventional evolution includes crossover, mutation and selection operations.

Crossover: As shown in Fig.9, for both individuals, one breakpoint is randomly determined and then their contents behind the breakpoint in both individuals are exchanged to build two new ones.

Mutation: For one individual as shown in Fig.8, randomly choose one element except peak bin value $P_k, k \in [1, q]$ in it. Then replace it with another value. For example, If one zero bin Z_k is chosen, another unused zero bin from **Zero_Set** could be used to replace current one. If the number of reuse or pattern number is chosen, the similar operations could be performed.

Selection: Roulette wheel selection [30] is employed to select those one with less distortion for crossover and mutation operations to reproduce the next generation.

2) EMPIRICAL MODIFICATION

The empirical modification for one individual could be performed towards two different directions, i.e., increasing embedding capacity or reduction of it. If the embedding capacity of current individual is larger than given payload, the modification towards the direction of reducing the embedding capacity is performed in order to make our solution best match the given payload and meanwhile acquire the least distortion. Otherwise, the modification towards the direction of increasing the embedding capacity.

The empirical modifications towards both directions are similar. Now, we take the modification towards the direction of reducing embedding capacity as an example to describe the operation as follows.

In the actual empirical modification process, one could be randomly utilized.

(1) Randomly reduce the number of reuse RT_P_k for one peak bin P_k in current individual unless $RT\_P_k = 0$;

(2) Choose a better pattern to replace the current one unless the current pattern is the best one. For example, as mentioned in Fig.6, Pattern 2 is better than Pattern 1.

D. PERFORMANCE EVALUATION

For each individual associated with those parameters as shown in Fig.8, the rate and distortion performance evaluation as mentioned in Section IV is implemented to rapid evaluate its rate (embedding capacity) and distortion D.

VI. EMBEDDING AND EXTRACTION PROCESSES

In general, the aim of proposed scheme is to determine the unconstraint optimal side information, i.e., peak and zeros bins, for ''multilevel embedding''. Then ''multilevel embedding'' process is performed for data hiding, which is fit for different carriers, either pixel or prediction error domain.

In our paper, similar embedding framework as [30] is employed, which includes: (1) The accurate rhombus prediction proposed in [19] is incorporated to generate prediction errors; (2) Prediction errors are sorted in ascending order based on their local variances [19] and then the top $\lambda\%$ of prediction errors corresponding to the smooth region are taken for HS based multilevel embedding. The optimal $\lambda\%$ is determined by given payload and described in [30].

A. EMBEDDING PROCESS

For a 8-bit grayscale cover image and the binary secret message $\mathbf{w}$ of length C, the embedding process is briefly described as follows.

1) INITIALIZATION

Divide the cover image in Cross (X) and Round (O) sets as shown in Fig.10(a), and assign half of the secret message for Cross and Round sets embedding, respectively.

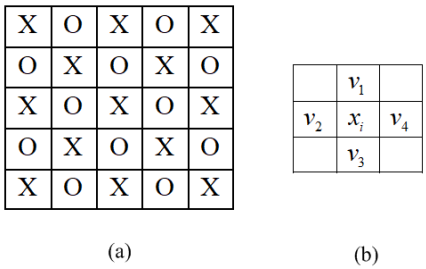


FIGURE 10. The sketch of rhombus prediction.

2) AREAS RESERVATION FOR AUXILIARY INFORMATION

Reserve the first two rows and columns in the cover image virgin to hide auxiliary information of "Round" set, denoted as $Aux(O)$, for reversible restoration.

$Aux(O)$ mainly includes chosen peak and zero bins, optimal $\lambda\%$ value and location map to avoid overflow/undeflow for "Round" set. Among them, location map generation as a common technique in RDH could be performed by many previous schemes, such as literature [22] employed in our paper.

3) RHOMBUS PREDICTION OF THE CROSS SET

As shown in Fig.10(b), each pixel x_i in the Cross set is predicted by its four neighboring pixels $v_i, i \in [1, 4]$ in Round set to obtain the prediction value $\widehat{x_i}$ by

$$\widehat{x_i} = \left\lceil \frac{v_1 + v_2 + v_3 + v_4}{4} \right\rceil \tag{16}$$

where $\lceil \bullet \rceil$ is ceiling function returning the nearest integer no less than the input.

Then the prediction error e_i is computed by

$$e_i = x_i - \widehat{x_i} \tag{17}$$

4) DETERMINE THE OPTIMAL $\lambda\%$ OF PREDICTION ERRORS IN THE CROSS SET TO BUILD CARRIER FOR DATA HIDING

Based on the histogram of prediction errors in the Cross set, i.e., $\mathbf{PE} = e_i | i \in [1, size]$ and assigned half of the secret message, determine the optimal top $\lambda\%$ of prediction errors as mentioned in [30] to build sub-histogram, denoted as **Sub_PE**, as carrier for data hiding.

5) OPTIMAL PEAK AND ZERO BINS SELECTION

Based on the generated **Sub_PE** in the Cross set, GA based scheme as mentioned in Section V is carried out to determine the optimal peak and zero bins for multilevel embedding.

6) CROSS SET EMBEDDING

To hide the message into **Sub_PE**, the accumulated shifts Δbin_i for each bin after the multilevel embedding are calculated by (8), (14) and (15) to generate the marked prediction errors in Cross set as follows.

Let e_i and $\widetilde{e_i}$ mean an original prediction error and its marked version, respectively. Then some cases should be individually performed as follows.

If i-th bin belongs to no peak bin, no secret data could be embedded and then

$$\tilde{e}_i = e_i + \Delta bin_i, i \in \underline{non-peakbin} \tag{18}$$

If i-th bin belongs to single utilized peak bin, 1- bit secret data $w = 0\text{or}1$ could be hidden and then

$$\tilde{e}_i = \begin{cases} e_i + \Delta bin_i + w, \\ e_i + \Delta bin_i - w, \end{cases} i \in \underline{single-peakbin} \tag{19}$$

Otherwise, if i-th bin belongs to repeatedly utilized peak bin, several bits secret data $\mathbf{w}$ could be hidden and then the embedded processing should be performed according to the employed pattern number. We take Fig.6 as an example. Assume Pattern 1 is used, then

$$\tilde{e}_i = \begin{cases} e_i + \Delta bin_i, whenw = 00, \\ e_i + \Delta bin_i + 1, whenw = 01, \\ e_i + \Delta bin_i + 2, whenw = 1, \end{cases} i \in \underline{repeated-peakbin} \tag{20}$$

7) STEGO-PIXELS GENERATION IN THE CROSS SET

With marked prediction errors, we have the stego-pixels in the cross set as follows.

$$\widetilde{x_i} = \widetilde{e_i} + \widehat{x_i} \tag{21}$$

During above process, those auxiliary information for "Cross" set, denoted as $Aux(X)$, are generated and will be hidden into the "Round" set.

8) STEGO-PIXELS GENERATION IN THE ROUND SET

Utilize the generated stego-pixels in the Cross set to similarly predict the pixels by using Step 3). Then hide $Aux(X)$, the LSBs of Round set in reserved areas mentioned in Step 2), and the rest half of secret data into Round set in order by repeating Step 4)-7).

TABLE 1. Comparison of two GA based scheme with or without Transfer Learning technique (TL) for test image Lena.

Bpp	0.05	0.3	0.5	0.7	1.0	1.2
With Transfer Learning (TL)	59.280	47.584	43.001	39.680	34.188	31.286
Without Transfer Learning (TL)	59.081	46.829	42.320	39.127	33.850	–

9) STEGO IMAGE GENERATION

Combine the stego-pixels in both Cross and Round sets, and then embed the $Aux(O)$ into the Round set of reserved region by using LSB replacement to generate stego-image.

Note that, in this paper, those auxiliary information for "Round" sets $Aux(O)$ is saved in the reserved area in order to ensure that it could be acquired by receiver and thus guide data extraction and cover image restoration.

B. EXTRACTION PROCESS

Firstly, extract the auxiliary information $Aux(O)$ from the reserved area of the stego-image. Then, based on stego-image and received optimum $\lambda\%$, peak and zero bins from $Aux(O)$, we extract half of the secret message and recover the "Round" set in the inverse order as embedding process. Finally, it is used to reconstruct the original Cross set and extract the rest hidden message.

VII. EXPERIMENTAL RESULTS

To evaluate the performance of our proposed scheme, we test six typical $512 \times 512 \times 8$ bits grayscale images with different texture characteristics from SIPI database [44]. Meanwhile, the effect of Transfer Learning technique involved in our scheme is also discussed as follows.

A. THE EFFECTIVENESS OF TRANSFER LEARNING TECHNIQUE (TL)

In the subsection, the effectiveness of Transfer Learning technique (TL) as shown in Fig.7 is testified. As listed in Table 1, PSNRs (Peak Signal to Noise Ratio) of the stego-image based on different schemes are offered for test image Lena. It is observed that the performances with TL is much better than those ones without TL at various payloads. The reason is that it is not easy to search desired side information, i.e., the combination of peak and zero bins for given payload, around such an entire huge solution space, which might lead to unstable results. With the help of Transfer Learning, some high-performance initial individuals (seeds), i.e., our previous good seed [30] from Level 2, could be introduced in our framework as shown in Fig.7 to guide a better search. In addition, since the previous work [30] is performed in relatively smaller solution space, its stability and high efficiency are guaranteed. It is noted that, to control the imperceptible distortion of stego-image, when PSNR for stego-image is less than 30, the embedding process is stopped, which is denoted as notation '–' in Table 1.

B. THE EFFECTIVENESS OF REPEATEDLY UTILIZED PEAK BINS IN MULTILEVEL EMBEDDING

To testify the effectiveness of the "multilevel embedding", comparison between "multilevel embedding" and "multiple embedding" are performed, where the former one allows to repeatedly utilize peak bins, while the latter is just associated with single utilized peak bins as illustrated in Fig.5, respectively. To make a fair comparison, the cover image is identically generated based on the rhombus prediction [19]. The results for three typical test images are listed in Table 2, where notation '–' means the payload could not be embedded or the quality of steo-image is too low (PSNR < 30). It is observed that: (1) For the small to medium payload, only single utilization of those peak bins located in smooth area is enough, and thus "multiple embedding" and "multilevel embedding" almost achieve the comparable performance. (2) With the increase of given payload, rather than employment of some peak bins in texture regions or exhausting all the possible desired regions with single utilization, repeated use of those peak bins located at smooth area might be a better choice, which demonstrates the superiority of multilevel embedding.

Table 3 lists some side information (namely those chosen peak and zero bins) at different payloads for test image F16, where the notation $[P_k, Z_k]$ is utilized to denote the case of single utilized peak bin with its corresponding zero bin in RDH process, such as $[-1, -6]$. While $[-1(2, 2), \{-23, -26\}]$ as shown in Table 3 is employed to denote the case of repeatedly utilized peak bin associated with its corresponding zero bins, where $P_k = -1$ and some information in the parenthesis indicates that the number of reuse $RT\_P_k = 2$ and associated pattern number $PN\_P_k = 2$ for multilevel embedding as mentioned in Fig.8. In addition, corresponding two zero bins are listed in the brace.

According to above analysis, it is concluded that "multilevel embedding" could achieve a better performance at different kinds of payload (from small to large payload),

TABLE 2. Comparison of two different embedding schemes ("multilevel embedding" and "multiple embedding").

Test images	Schemes	Given Payload (Unit: BPP)					
		0.05	0.3	0.5	0.7	1.0	1.2
Lena	Multilevel Embedding	59.280	47.633	43.046	39.695	34.188	31.286
	Multiple Embedding	59.182	47.583	43.001	39.680	–	–
Baboon	Multilevel Embedding	53.664	37.912	32.609	27.928	–	–
	Multiple Embedding	53.598	37.884	32.555	–	–	–
F16	Multilevel Embedding	62.202	51.514	46.933	43.081	36.063	33.276
	Multiple Embedding	62.098	51.465	46.888	43.039	–	–

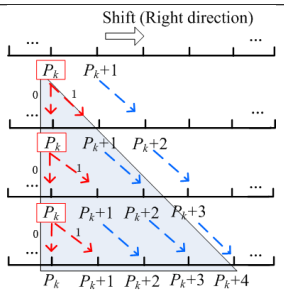
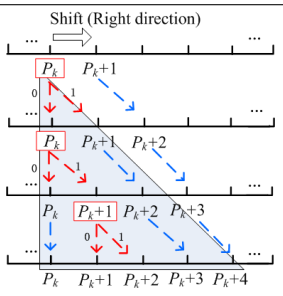
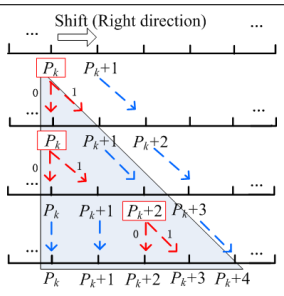
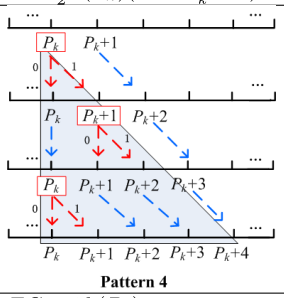
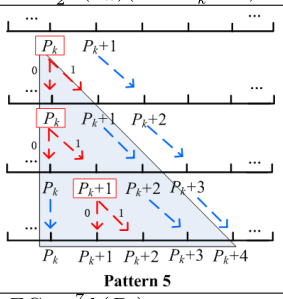
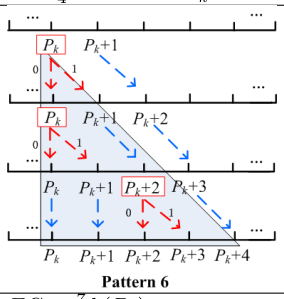
TABLE 3. The detailed side information for different embedding schemes ("multilevel embedding" and "multiple embedding") based on test image F16.

Bpp	Schemes	PSNR	Side information			
			X-set		O-set	
			Optimal λ%	$[P_k, Z_k]$	Optimal λ%	$[P_k, Z_k]$
0.05	"multilevel embedding"	62.23	12%	[-1, -6], [1 , 5]	19%	[-2 , -8], [1 , 7]
	"multiple embedding"	62.03	12%	[-1, -6], [1, 5]	23%	[-3 , -8], [1 , 7]
0.8	"multilevel embedding"	41.03	86%	[-11, -37], [-10, -35], [-6, -33], [-4, -32], [-2, -29], **[-1 (2, 2), {-23, -26}]**, [0, 28], [1, 29], [2, 30], [3, 31], [6, 32]	86%	[-6, -31], [-5, -30], [-4, -29], [-3, -27], [-2, -26], [-1, -25], [0, 23], [1, 26], [2, 27], [3, 28], [4, 29], [5, 31], [6, 32], [10, 35]
	"multiple embedding"	40.88	85%	[0, -23], [1, 23], [5, -26], [-5, -29], [4, 28], [-4, -95], [3, 29], [-3, 43], [2, 41], [-2, -128], [-1, -37], [6, 32], [-14, 229], [-12, -127], [-9, -66]	88%	[0, -36], [-1, -14], [5, 38], [-5, -30], [4, 29], [-4, -31], [3, 31], [-3, -32], [2, -52], [-2, 42], [1, -53], [-10, 43], [-7, 13]

TABLE 4. Computation time (CmpTime) comparison between different embedding schemes based on different test images at Bpp = 0.05 (Unit:Second).

Schemes	Test images						Average
	Lena	Baboon	F16	Peppers	Boat	Barbara	
Multilevel Embedding	169.79	192.03	188.72	173.40	171.57	171.39	**177.82**
Multiple Embedding	34.63	61.23	44.73	37.65	44.79	37.03	**43.34**

TABLE 5. The patterns for peak bin P_k associated with repeatedly times $RT\_P_k = 3$.

Multilevel embedding	 Pattern 1	 Pattern 2	 Pattern 3
Performance	$EC = \frac{7}{4}h(P_k)$ $D_{repeated-peak} =$ $\frac{1}{8}h(P_k)(\Delta bin_{P_k}+1)^2$ $+\frac{1}{4}h(P_k)(\Delta bin_{P_k}+2)^2$ $+\frac{1}{2}h(P_k)(\Delta bin_{P_k}+3)^2$	$EC = \frac{7}{4}h(P_k)$ $D_{repeated-peak} =$ $\frac{1}{8}h(P_k)(\Delta bin_{P_k}+1)^2$ $+\frac{1}{4}h(P_k)(\Delta bin_{P_k}+2)^2$ $+\frac{1}{2}h(P_k)(\Delta bin_{P_k}+3)^2$	$EC = 2h(P_k)$ $D_{repeated-peak} =$ $\frac{1}{4}h(P_k)(\Delta bin_{P_k}+1)^2$ $+\frac{1}{4}h(P_k)(\Delta bin_{P_k}+2)^2$ $+\frac{1}{4}h(P_k)(\Delta bin_{P_k}+3)^2$
Multilevel embedding	 Pattern 4	 Pattern 5	 Pattern 6
Performance	$EC = 2h(P_k)$ $D_{repeated-peak} =$ $\frac{1}{4}h(P_k)(\Delta bin_{P_k}+1)^2$ $+\frac{1}{4}h(P_k)(\Delta bin_{P_k}+2)^2$ $+\frac{1}{4}h(P_k)(\Delta bin_{P_k}+3)^2$	$EC = \frac{7}{4}h(P_k)$ $D_{repeated-peak} =$ $\frac{1}{8}h(P_k)(\Delta bin_{P_k}+1)^2$ $+\frac{1}{4}h(P_k)(\Delta bin_{P_k}+2)^2$ $+\frac{1}{2}h(P_k)(\Delta bin_{P_k}+3)^2$	$EC = \frac{7}{4}h(P_k)$ $D_{repeated-peak} =$ $\frac{1}{4}h(P_k)(\Delta bin_{P_k}+1)^2$ $+\frac{1}{4}h(P_k)(\Delta bin_{P_k}+2)^2$ $+\frac{1}{4}h(P_k)(\Delta bin_{P_k}+3)^2$

which attributes to the unconstraint optimal side information selection mechanism.

C. COMPREHENSIVE PERFORMANCE COMPARISON WITH OTHER SCHEMES

Finally, comprehensive comparison between our scheme and other state-of-the-art algorithms, such as Sachnev *et al.* [19], Luo *et al.* [22], Wang *et al.* [30] and so on, is implemented. The results are offered in Fig.11. It is observed that proposed scheme almost achieves the best performance in the entire range, i.e., from small payload to large one. And its embedding capacity could be up to more than 1.2bpp (bit per pixel) for some test images. The superiority results from two reasons: (1) The mechanism of unconstraint optimal side

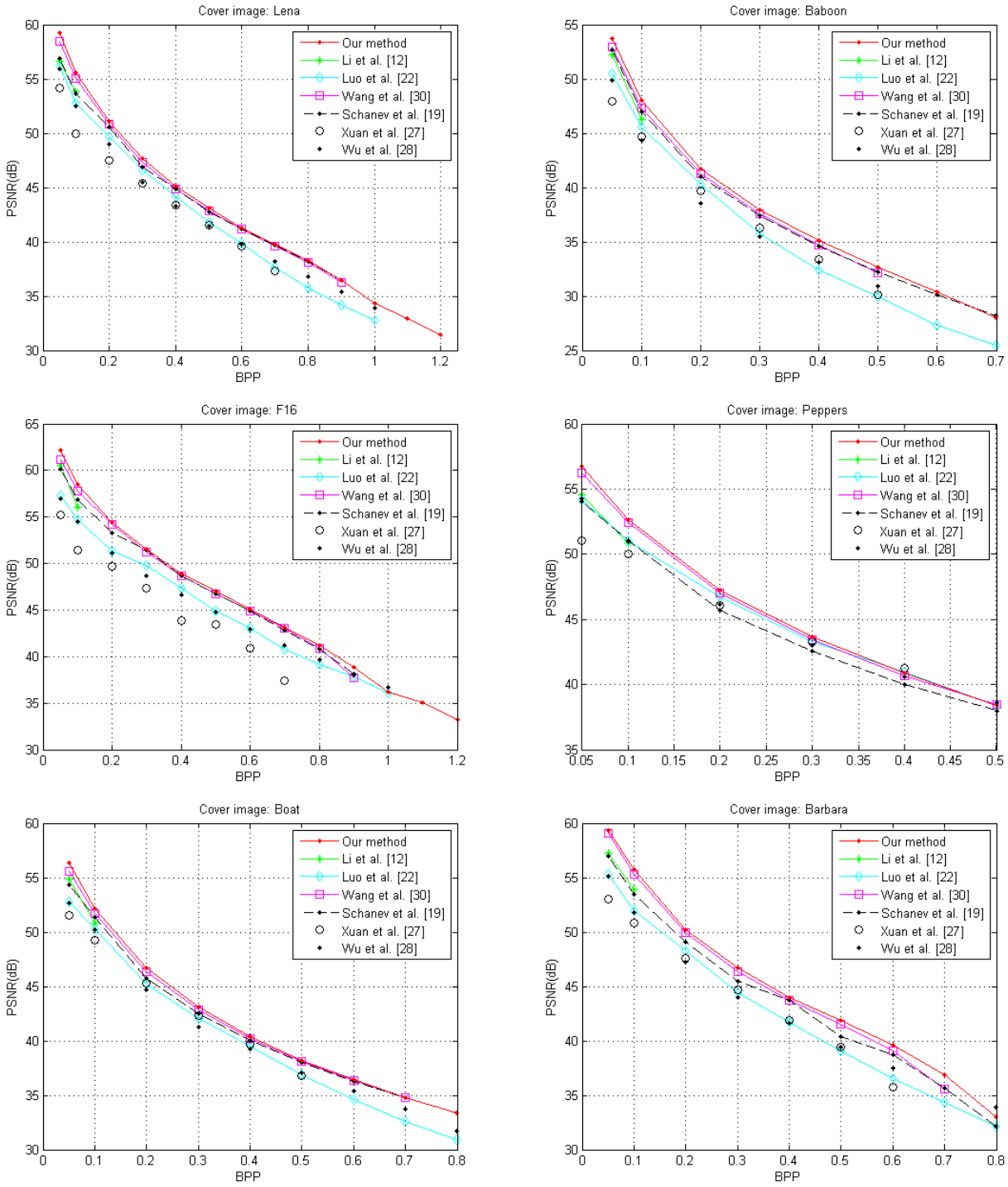


FIGURE 11. Comprehensive performance comparison.

information selection ensures the essentially global optimal side information. (2) Some specially designed algorithms, i.e., transfer learning based GA scheme, further guarantee a rapid and stable search towards the desired search direction.

D. THE COMPUTATION COST

In this subsection, we evaluate the practical computation time (CmpTime) of both ''multilevel embedding'' and ''multiple embedding'' schemes for different test images at 0.05bpp, respectively. The simulation platform is MATLAB based on a 2.7 GHz Intel Pentium Dual Core CPU with 8GB memory. The results are listed in Table 4. Evidently, the computation times for our multilevel embedding scheme are relatively larger compared with ''multiple embedding'' algorithm due to two factors: (1) Our scheme will perform much more flexible search among a larger solution space

associated with complex ''multilevel embedding''; (2) Transfer Learning model involved in our scheme is actually a ''multiple embedding'' process and thus spends some computation time. However, due to some rapid algorithms specially designed in our paper to control the computation complexity as mentioned in Section VI and V, the computational cost for our scheme is affordable.

VIII. CONCLUSION

In this paper, an unconstraint optimal selection of side information, i.e., peak and zero bins, for multilevel embedding based reversible data hiding is proposed. Due to flexible search of those side information around an entire huge solution space, expected performance is achieved at different payloads. In addition, to reduce the computation complexity for ''multilevel embedding'' and achieve stable search among the entire huge solution space, rapid performance evaluation method and transfer learning based GA search scheme are designed. Experimental results show the superiority of the proposed scheme and meanwhile achieve an affordable time cost. Finally, our proposed scheme will be extended to multiple histograms based RDH [31] and other cases in future.

IX. APPENDIX

As mentioned in Section IV, those patterns for peak bin P_k associated with the number of reuse $RT\_P_k = 3$ based on shifting towards ''right direction'' are pre-prepared and listed in Table 5.

REFERENCES

[1] D. Wu, F. Zhang, H. Wang, and R. Wang, ''Fundamental relationship between node dynamic and content cooperative transmission in mobile multimedia communications,'' *Comput. Commun.*, vol. 120, pp. 71–79, May 2018.

[2] F. Almasalha, F. Naït-Abdesselam, G. Trajcevski, and A. Khokhar, ''Secure transmission of multimedia contents over low-power mobile devices,'' *J. Inf. Secur. Appl.*, vol. 40, pp. 183–192, Jun. 2018.

[3] Z. Pan, X. Yi, and L. Chen, ''Motion and disparity vectors early determination for texture video in 3D-HEVC,'' *Multimedia Tools Appl.*, pp. 1–18, Nov. 2018.

[4] Z. Pan, J. Lei, Y. Zhang, and F. Wang, ''Adaptive fractional-pixel motion estimation skipped algorithm for efficient HEVC motion estimation,'' *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 14, no. 1, pp. 1–19, 2018.

[5] J. Lei, J. Duan, F. Wu, N. Ling, and C. Hou, ''Fast mode decision based on grayscale similarity and inter-view correlation for depth map coding in 3D-HEVC,'' *IEEE Trans. Circuits Syst. Video Technol*, vol. 28, no. 3, pp. 706–718, Mar. 2018.

[6] J. Fridrich, M. Goljan, and R. Du, ''Invertible authentication,'' *Proc. SPIE*, vol. 4314, pp. 197–208, Aug. 2001.

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

[8] J. Tian, ''Reversible watermarking using a difference expansion,'' *IEEE Trans. Circuits Syst. Video Technol*, vol. 13, no. 8, pp. 890–896, 2003.

[9] 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.

[10] M. Fallahpour, ''Reversible image data hiding based on gradient adjusted prediction,'' *IEICE Electron. Express*, vol. 5, no. 20, pp. 870–876, 2008.

[11] D. Coltuc, ''Improved embedding for prediction-based reversible watermarking,'' *IEEE Trans. Inf. Forensics Security*, vol. 6, no. 3, pp. 873–882, Sep. 2011.

[12] 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.

[13] 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.

[14] 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.

[15] I. C. Dragoi and D. Coltuc, ''On local prediction based reversible watermarking,'' *IEEE Trans. Image Process.*, vol. 24, no. 4, pp. 1244–1246, Apr. 2015.

[16] W. Hong, T.-S. Chen, and J. Chen, ''Reversible data hiding using Delaunay triangulation and selective embedment,'' *Inf. Sci.*, vol. 308, pp. 140–154, Jul. 2015.

[17] 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.

[18] W.-L. Tai, C.-M. Yeh, and C.-C. Chang, ''Reversible data hiding based on histogram modification of pixel differences,'' *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 6, pp. 906–910, Jun. 2009.

[19] 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.

[20] H. J. Hwang, H. J. Kim, V. Sachnev, and S. H. Joo, ''Reversible watermarking method using optimal histogram pair shifting based on prediction and sorting,'' *KSII Trans. Internet Inf. Syst.*, vol. 4, no. 4, pp. 655–670, 2010.

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

[22] 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.

[23] X. Li, B. Li, B. Yang, and T. Zeng, ''General framework to histogram-shifting-based reversible data hiding,'' *IEEE Trans. Image Process.*, vol. 22, no. 6, pp. 2181–2191, Jun. 2013.

[24] C. Qin, C.-C. Chang, Y.-H. Huang, and L.-T. Liao, ''An inpainting-assisted reversible steganographic scheme using a histogram shifting mechanism,'' *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 7, pp. 1109–1118, Jul. 2013.

[25] F. Peng, X. Li, and B. Yang, ''Improved PVO-based reversible data hiding,'' *Digit. Signal Process.*, vol. 25, no. 2, pp. 255–265, 2014.

[26] X. Qu and H. J. Kim, ''Pixel-based pixel value ordering predictor for high-fidelity reversible data hiding,'' *Signal Process*, vol. 111, no. 1, pp. 249–260, 2015.

[27] G. Xuan, X. Tong, J. Teng, X. Zhang, and Y. Q. Shi, ''Optimal histogram-pair and prediction-error based image reversible data hiding,'' in *Proc. Int. Workshop Digit.-Forensics Watermarking*, Shanghai, China, 2012, pp. 368–383.

[28] H.-T. Wu and J. Huang, ''Reversible image watermarking on prediction errors by efficient histogram modification,'' *Signal Process.*, vol. 92, no. 12, pp. 3000–3009, Dec. 2012.

[29] X. Ma, Z. Pan, S. Hu, and L. Wang, ''High-fidelity reversible data hiding scheme based on multi-predictor sorting and selecting mechanism,'' *J. Vis. Commun. Image Represent.*, vol. 28, pp. 71–82, Apr. 2015.

[30] J. Wang, J. Ni, X. Zhang, and Y.-Q. Shi, ''Rate and distortion optimization for reversible data hiding using multiple histogram shifting,'' *IEEE Trans. Cybern.*, vol. 47, no. 2, pp. 315–326, Feb. 2017.

[31] X. Li, W. Zhang, X. Gui, and B. Yang, ''Efficient reversible data hiding based on multiple histograms modification,'' *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 9, pp. 2016–2027, Sep. 2015.

[32] H. Chen, J. Ni, W. Hong, and T.-S. Chen, ''High-fidelity reversible data hiding using directionally enclosed prediction,'' *IEEE Signal Process. Lett.*, vol. 24, no. 5, pp. 574–578, May 2017.

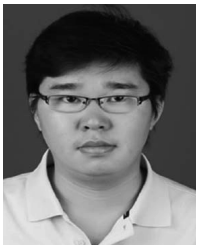
[33] B. Ou, X. Li, and J. Wang, ''Improved PVO-based reversible data hiding: A new implementation based on multiple histograms modification,'' *J. Vis. Commun. Image Represent.*, vol. 38, pp. 328–339, Jul. 2016.

[34] C.-C. Lin, W.-L. Tai, and C.-C. Chang, ''Multilevel reversible data hiding based on histogram modification of difference images,'' *Pattern Recognit.*, vol. 41, no. 12, pp. 3582–3591, 2008.

[35] X. Hu, W. Zhang, X. Li, and N. Yu, ''Minimum rate prediction and optimized histograms modification for reversible data hiding,'' *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 3, pp. 653–664, Mar. 2015.

[36] D. Coltuc, ''Low distortion transform for reversible watermarking,'' *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 412–417, Jan. 2012.

[37] F. Peng, X. Li, and B. Yang, "Adaptive reversible data hiding scheme based on integer transform," *Signal Process.*, vol. 92, no. 1, pp. 54–62, 2012.
[38] B. Ma and Y. Q. Shi, "A reversible data hiding scheme based on code division multiplexing," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 9, pp. 1914–1927, Sep. 2016.
[39] L. Xiang, Y. Li, W. Hao, P. Yang, and X. Shen, "Reversible natural language watermarking using synonym substitution and arithmetic coding," *Comput., Mater. Continua*, vol. 55, no. 3, pp. 541–559, 2018.
[40] Y. Chen, B. Yin, H. He, S. Yan, F. Chen, and H. Tai, "Reversible data hiding in classification-scrambling encrypted-image based on iterative recovery," *Comput., Mater. Continua*, vol. 56, no. 2, pp. 299–312, 2018.
[41] L. Xiong and Y. Shi, "On the privacy-preserving outsourcing scheme of reversible data hiding over encrypted image data in cloud computing," *Comput., Mater. Continua*, vol. 55, no. 3, pp. 523–539, 2018.
[42] Y.-Q. Shi, X. Li, X. Zhang, H.-T. Wu, and B. Ma, "Reversible data hiding: Advances in the past two decades," *IEEE Access*, vol. 4, pp. 3210–3237, 2016.
[43] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*. Reading, MA, USA: Addison-Wesley, 1989.
[44] [Online]. Available: http://sipi.usc.edu/database/



JUNXIANG WANG received the M.S. degree from the Harbin Institute of Technology, Harbin, China, in 2008, and the Ph.D. degree from Sun Yat-sen University, Guangdong, China, in 2012. He has been a Visiting Scholar with the New Jersey Institute of Technology, Newark, NJ, USA, since 2017.

He is currently an Associate Professor with the School of Mechanical and Electronic Engineering, Jingdezhen Ceramic Institute, Jiangxi, China. His current research interests include information security and image processing.

XIN CHEN received the B.S. degree from the Jingdezhen Ceramic Institute, Jiangxi, China, in 2017, where he is currently pursuing the M.S. degree with the School of Information Science and Technology. His current research interests include digital watermarking and image processing.

YUNQING SHI received the M.S. degree from Shanghai Jiao Tong University, Shanghai, China, and the Ph.D. degree from the University of Pittsburgh, Pittsburgh, PA, USA. He has been with the New Jersey Institute of Technology, Newark, NJ, USA, since 1987. He has authored or co-authored 300 papers, one book, five book chapters, and an Editor of ten books. He holds 28 U.S. patents. His current research interests include digital data hiding, forensics and information assurance, visual signal processing, and communications.

He is a member of a few IEEE technical committees. He was a recipient of the Innovators Award 2010 by New Jersey Inventors Hall of Fame for Innovations in Digital Forensics and Security and the 2010 Thomas Alva Edison Patent Award by Research and Development Council of New Jersey for the U.S. patent 7 457 341 entitled System and Method for Robust Reversible Data Hiding and Data Recovery in the Spatial Domain. He also served as a Technical Program Chair for the IEEE ICME07, a Co-Technical Chair for IWDW06–13 and the IEEE MMSP05, and a Co-General

Chair for the IEEE MMSP02. He served as an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING and the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS (II), and a few other journals. He also served as a Distinguished Lecturer for the IEEE CASS.