Minimum Rate Prediction and Optimized Histograms Modification for Reversible Data Hiding

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

***Abstract*— Prediction-error expansion (PEE)-based reversible data hiding schemes consist of two steps. First, a sharp prediction-error (PE) histogram is generated by utilizing pixel prediction strategies. Second, secret messages are reversibly embedded into the prediction-errors through expanding and shifting the PE histogram. Previous PEE methods treat the two steps independently while they either focus on pixel prediction to obtain a sharp PE histogram, or aim at histogram modification to enhance the embedding performance for a given PE histogram. This paper propose a pixel prediction method based on the minimum rate criterion for reversible data hiding, which establishes the consistency between the two steps in essence. And correspondingly, a novel optimized histograms modification scheme is presented to approximate the optimal embedding performance on the generated PE sequence. Experiments demonstrate that the proposed method outperforms the previous state-of-art counterparts significantly in terms of both the prediction accuracy and the final embedding performance.**

***Index Terms*— Prediction-error expansion, reversible data hiding, minimum rate prediction, optimized histograms modification.**

I. Introduction

REVERSIBLE data hiding (RDH) [1] is a special embedding technique that ensures not only the embedded messages be extracted precisely, but also the cover itself should be restored losslessly. This important technique is widely used in medical imagery [2], military imagery [3] and law forensics, where the original cover is too precious to be damaged. Moreover, it has been found that RDH can be quite helpful in video error-concealment coding [4].

A plenty of RDH algorithms have been proposed in the past decade. Until now, they roughly fall into three categories: the compression appending framework [3], the difference expansion (DE) [5], [6], [8] scheme and the histogram shift (HS) technique [9]. In fact, by applying DE or HS to the residual parts of images instead, e.g., the prediction errors (PE) [10]–[14], better performance can be achieved. This extended method is called prediction-error expansion (PEE), which is currently a research hot spot and the most powerful technique of RDH.

Almost all state-of-art PEE algorithms consist of two steps. The first step generates a PE sequence with small entropy, i.e., the PE has a sharp histogram which usually can be realized by utilizing kinds of pixel prediction strategies combined with the sorting [11] or pixel selection [14] technique. The second step reversibly embeds the message into the PE sequence by modifying its histogram with methods like HS and DE.

As for the second step, the upper bound of the payload for a fixed host sequence and a distortion constraint is given by Kalker and Willems [15]. They formulated the RDH as a special rate-distortion problem, and obtained the rate-distortion function, i.e., the upper bound of the embedding rate under a given distortion constraint Δ, as follows:

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

where X and Y denote the random variables of 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 $\sum_{x,y} P_X(x)P_{Y|X}(y|x)D(x,y) \leq \Delta$. The distortion metric $D(x,y)$ is usually defined as the square error distortion, $D(x,y) = (x-y)^2$.

Therefore, to evaluate the capacity of RDH given the host sequence, one can calculate the optimal transition probability matrix $P_{Y|X}(y|x)$ and then use it for guidance to approximate the upper bound of the embedding capacity. Lin et al. [16] proposed a method to estimate the optimal transition probability $P_{Y|X}(y|x)$, by which they can evaluate the capacity (1) for distortion metrics such as square error distortion or L_1-Norm distortion. In [17], we proposed a fast algorithm to estimate the optimal transition probability $P_{Y|X}(y|x)$ for both the distortion constrained problem (1) and its dual problem, i.e., the embedding rate constrained problem. Afterwards, by improving the recursive code construction (RCC), we obtain the optimal embedding method of RDH for binary host sequences [18], [19] and general gray-scale host sequences [20] respectively. Meanwhile we proved that RCC can approach the rate-distortion bound (1) as long as the entropy encoder reaches entropy, which establishes the equivalency between RDH and lossless data compression.

Manuscript received September 2, 2014; revised January 3, 2015; accepted January 5, 2015. Date of publication January 14, 2015; date of current version February 9, 2015. This work was supported in part by the Natural Science Foundation of China under Grant 61170234 and Grant 60803155, in part by the Strategic Priority Research Program through the Chinese Academy of Sciences under Grant XDA06030601, and in part by the Science and Technology on Information Assurance Laboratory under Grant KJ-13-003. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Patrick Bas.

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

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

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

Digital Object Identifier 10.1109/TIFS.2015.2392556

Thus to some extent, for the second step of PEE based RDH schemes, i.e., the capacity approaching embedding for a given PE sequence can be practically realized as discussed above. But for the first step of PEE schemes, here comes one nature problem. What is the optimal evaluation criteria for the goodness of a prediction strategy?

Typical prediction methods either use a fixed neighborhood average model [11], or a content adaptive predictor such as the median edge detector (MED) [6], [7], the gradient adjusted predictor (GAP) [23], [24] and the checker-board based predictor (CBP) [22]. Minimum Mean Square Error (MMSE) predictors are also designed to predict pixel values for PEE schemes like the global optimal weight-based predictor [25], or the recently proposed local-prediction-based least square (LS) predictor [26], in which a distinct LS predictor is calculated on a neighboring block for each pixel individually.

Either fixed predictors, or content adaptive predictors, or the LS predictors are aimed at generating a sharp PE histogram which is generally like a Laplace distribution. A sharp histogram means a small entropy for the PE sequence, i.e., a small amount of information bits required to encode the PE sequence. Observing from the rate-distortion formulation (1), a small value of $H(X)$ means more capacity exists in the difference between the two entropies $H(Y) - H(X)$, and thus better embedding performance can be achieved. But intrinsically, these above pixel predictors expect to fit the PE histogram to a single distribution like the Laplace or the Gaussian distribution, which to a certain extent restricts themselves from decreasing the total information bits to encode the PE sequence.

In [28], Matsuda et al. designed an adaptive pixel predictor for lossless image coding based on the minimum rate criterion, in which they aim to fit the prediction errors to a group of fixed Gaussian probability models, and then a context-adaptive arithmetic coding scheme is applied to encode the prediction errors. Recently, Gui et al. [27] proposed a high capacity adaptive embedding scheme based on generalized prediction-error expansion where they classify the pixels into several levels according to their neighborhood complexity measurements, and then different amount of data bits are assigned to different levels of pixels to be embedded. A rather similar idea to split histogram and adaptively choosing the bins of the histogram according to the embedding rate is introduced recently by Caciula et al. [29].

In this paper, we propose a pixel prediction strategy for PEE based reversible data hiding schemes based on the minimum rate criterion, which was inspired by the lossless image coding methods in [28] and the context modeling of JPEG-LS standard [7]. Instead of pursuing a single sharp PE histogram, we design the pixel predictors for the sake of minimizing the conditional entropy of the PE sequence by utilizing mixture of Gaussian distributions. And correspondingly, a novel optimized multiple histograms modification scheme is presented to finally embed messages into the generated Gaussian mixture of PE sequence, which automatically allocates different amount of data into different groups of pixels like the schemes in [27].

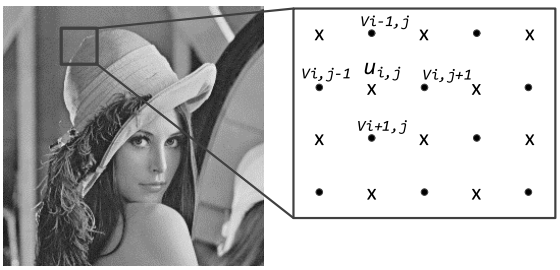


Fig. 1. Rhombus prediction pattern. The pixel value of u of the Cross set is predicted by using its four neighboring pixels in the Dot set.

The rest of the paper is structured as follows. Section II-A briefly introduces the PEE based reversible data hiding schemes and the rate-distortion bound approaching embedding realizations. The proposed minimum rate prediction method is presented in Section III, followed by the novel optimized histograms modification strategy in Section IV. Experimental comparison results are demonstrated in Section V. And finally, some discussions and conclusions are given in Section VI.

II. Prediction-Error Expansion and Optimized Histogram Modification

A. Prediction-Error Expansion

Typical PEE based schemes divide cover image pixels into different parts, while a pixel of one part is predicted by its neighboring pixels in other parts. For the rhombus prediction pattern in Sachnev et al.'s double-layered embedding method [11], all pixels are divided into two sets: the Cross set and the Dot set (see Fig. 1). In the first round, the Cross set is used for embedding data and Dot set for computing predictions, while in the second round, the Dot set is used for embedding and Cross set for computing predictions. Since the two layers' embedding processes are similar in nature, we only take the Cross layer for illustration.

As shown in Fig. 1, the Cross pixels $u_{i,j}$ in the cover image are collected into a sequence $\mathbf{u} = (u_1, u_2, \cdots, u_n)$ from left to right and from top to bottom. For each Cross pixel $u_{i,j}$, the rhombus predicted value $\hat{u}_{i,j}$ is computed by averaging its four nearest Dot pixels:

$$\hat{u}_{i,j} = \lfloor \frac{v_{i,j-1} + v_{i+1,j} + v_{i,j+1} + v_{i-1,j}}{4} \rfloor \qquad (2)$$

Then, by subtracting the predicted value $\hat{u}_{i,j}$ from the original pixel value $u_{i,j}$, we obtain the prediction-error sequence $\mathbf{e} = (e_1, e_2, \cdots, e_n)$. Afterwards, secret data are embedded into the prediction-error sequence $\mathbf{e}$ through expanding and shifting techniques. Details of the embedding and extraction procedures for the standard PEE scheme are illustrated in [11].

B. Optimized Histogram Modification

Actually, for a given PE sequence, the upper bound of the embedding rate under an input distortion constraint Δ is given by (1). So instead of using the expanding and shifting technique described in [11], Lin et al. [16] proposed a pixel by pixel code construction to approach the rate distortion bounds for distortion metrics like square error distortion or L_1-Norm distortion. And alternatively, by improving the recursive code

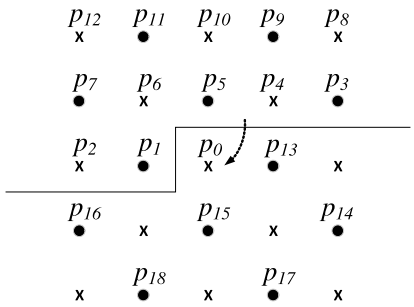


Fig. 2. Prediction context of a Cross set pixel.

construction (RCC), we obtain the optimal embedding method for general gray-scale PE sequences [20], which performs in a bin by bin manner respectively.

In [16], Lin et al.'s optimal code construction needs the estimated optimal marginal distribution $P_Y(y)$ of the marked signal as input, while in [17], we proposed a fast algorithm to estimate the optimal marginal distribution $P_Y(y)$ for both distortion constrained and rate constrained problems. Therefore, to realize the optimal code construction practically, one can first estimate the optimal marginal distribution $P_Y(y)$ of the marked signal, and then use Lin et al.'s code construction to embed messages. For the rest of the paper, we name this combined scheme as the optimized histogram modification (OHM), in which it treats the input PE sequence as a single probability distribution histogram.

Next we elaborate the OHM procedure to embed messages into the Cross set pixels as shown in Fig. 1.

- Firstly, using the rhombus prediction strategy in Section II-A, a prediction-error sequence $\mathbf{e} = (e_1, e_2, \cdots, e_n)$ is obtained from the Cross set pixels, which acts as the host signal X in (1).
- Secondly, in order to minimize the embedding distortion for a fixed length of message payload, run the fast algorithm in [17] to estimate the optimal marginal distribution $P_Y(y)$ of the marked signal Y, i.e. the marked prediction-error sequence $e' = (e'_1, e'_2, \cdots, e'_n)$.
- Finally, according to the estimated marginal distribution $P_Y(y)$, by modifying the host PE sequence $\mathbf{e}$ to the marked PE sequence $\mathbf{e}' = (e'_1, e'_2, \cdots, e'_n)$, messages can be optimally embedded through Lin et al's code construction [16].

Another OHM scheme which behaves in a bin by bin manner instead of Lin et al.'s pixel by pixel manner can be found in our recent work [20].

III. Minimum-Rate Prediction

In [28], Matsuda et al. designed an adaptive pixel predictor for lossless image coding based on the minimum rate criterion. Now we adapt the minimum rate criterion to optimize pixel predictors for PEE based reversible data hiding schemes. Firstly, like the scheme in Sachnev et al.'s method [11], pixels in image $\mathbf{I}$ are divided into two sets, the Cross set and the Dot set. Then for each pixel p_0 in the cover image, instead of predicting by averaging its four nearest neighbors in Fig. 1, we adopt the linear combinations of its K neighboring pixels for prediction. As shown in Fig. 2, the prediction error e at the current Cross set pixel p_0 is given by:

$$e = p_0 - \sum_{k=1}^{K} a_k \cdot p_k \tag{3}$$

where $p_k (k = 0, 1, \cdots, K)$ are pixel values of the image $\mathbf{I}$ and $a_k (k = 1, 2, \cdots, K)$ are the prediction coefficients. K is the prediction order. The prediction coefficients a_k are optimized later for the image and the values of them are coded as overhead information to be transmitted to the receiver side.

Assume that the probability distribution $P(e)$ of the prediction error e follows the simple zero mean Laplace form

$$P(e) = \frac{1}{2b} \exp\left(-\frac{|e|}{b}\right) \tag{4}$$

where b is the scale parameter.

From the information theory perspective, the average bits $L(e)$ required for encoding a value of e is given by

$$\begin{aligned} L(e) &= -\log_2(P(e)) \\ &= -\log_2(\frac{1}{2b}) + \log_2 \epsilon \cdot \frac{|e|}{b} \end{aligned} \tag{5}$$

while ϵ is the base of the natural logarithm. Consequently, considering all the Cross set pixels in a certain image region $\mathbf{R}$ of the image $\mathbf{I}$, the total average bits on the prediction errors can be estimated as:

$$\begin{aligned} J(\mathbf{R}) &= \sum_{p_0 \in \mathbf{R}} L(e) \\ &= -\sum_{p_0 \in \mathbf{R}} \log_2(\frac{1}{2b}) + \sum_{p_0 \in \mathbf{R}} \log_2 \epsilon \cdot \frac{|e|}{b} \end{aligned} \tag{6}$$

As a result, if we treat the whole image $\mathbf{I}$ as a large region, the average information cost will be $J(\mathbf{I})$.

Observing from (6), if the Laplace scale b is constant in the certain image region $\mathbf{R}$, the total information cost $J(\mathbf{R})$ will be determined by the sum of absolute of the prediction errors because the first term of $J(\mathbf{R})$ is a constant. Thus in this case, the predictor which minimizes the information cost function $J(\mathbf{R})$ reduces to the $L1$-norm minimization solution, which is similar to the least square (LS) predictors designed for reversible data hiding [25], [26]. However, natural images often contain kinds of textures and structures. Pixels in smooth places are easy to predict while in complex places pixel prediction becomes difficult. Thus the value of b should vary from a small one to a big one, with respect to the local contexts of different pixels. In order to chase the minimum rate predictor for nature images, it is necessary to assign a context adaptive value of b for each pixel individually.

A. Mixture of Probability Models

For all the Cross set pixels in a certain image region $\mathbf{R}$, instead of considering a single probability pattern, we assume that the probability density function (PDF) of the prediction error e can be modeled by the following generalized Gaussian mixture:

$$P_n(e) = \alpha_n \cdot \exp\left\{-|\eta(s_n, \sigma_n) \cdot e|^{s_n}\right\} \tag{7}$$

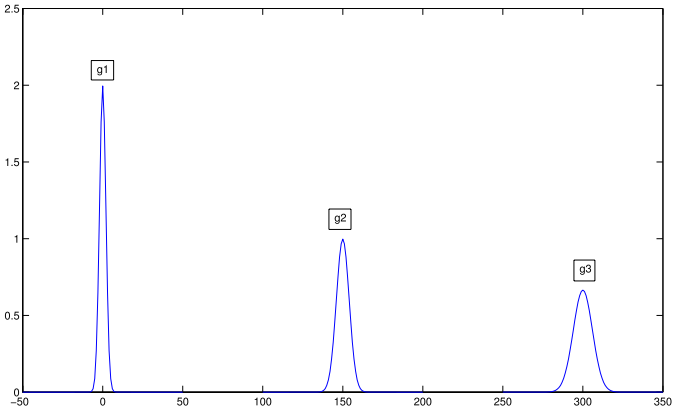


Fig. 3. Probability models of the prediction error e.

where α_n is a normalizing factor which makes the probabilities of all candidate prediction errors e sum to one. And the function $\eta(s_n, \sigma_n)$ is given by

$$\eta(s_n, \sigma_n) = \frac{1}{\sigma_n}\sqrt{\frac{\Gamma(3/s_n)}{\Gamma(1/s_n)}} \tag{8}$$

Here $\Gamma(\cdot)$ is the gamma function, σ_n is the standard deviation of e and s_n is a shape parameter which controls the sharpness of the PDF.

The region $\mathbf{R}$ consists of N probability model groups $\mathbf{g}_n(n = 1, 2, \cdots, N)$ and the values of σ_n and s_n in these groups differ from one to another, which controls the shape of the probability models as depicted in Fig. 3.

For the sake of conducting the group classification inside each image region $\mathbf{R}$, for each current pixel p_0, a local adaptive context parameter G is calculate as:

$$G = \sum_{k=2,4,6,8,10,12} \frac{1}{d_k} \cdot |e_k| \tag{9}$$

where d_k is the Euclidean distance of the neighboring pixel p_k from the current pixel p_0 as shown in Fig. 2, and e_k is the prediction error at the neighboring pixel p_k. Roughly speaking, the local adaptive parameter G measures the smoothness of the neighboring context for the current pixel p_0. So a small value of G indicates a smooth neighborhood, and a large value of G indicates a complex neighborhood like texture places or structural places. Therefore if the current pixel p_0 locates in a smooth context, mean a small value of G for instance, we can expect that the value of σ for the prediction error e at pixel p_0 will also be small, and vice versa.

Afterwards, by comparing the value of the parameter G with a set of thresholds $\{T_1 \le T_2 \le \cdots \le T_{N-1}\}$, we classify the current pixel p_0 into one of the N groups $\{\mathbf{g}_1, \mathbf{g}_2, \cdots, \mathbf{g}_N\}$. Specifically, if the parameter G of the current pixel p_0 satisfies the condition $T_{n-1} \le G < T_n (T_0 = 0, T_N = \infty)$, then we decide that it belongs to the group $\mathbf{g}_n$.

According to (7), the amount of information $L(e)$ required for encoding a value of e can be now rewritten as:

$$\begin{aligned} L_n(e) &= -\log_2(P_n(e)) \\ &= -\log_2 \alpha_n + \log_2 \epsilon \cdot |\frac{e}{\sigma_n}\sqrt{\frac{\Gamma(3/s_n)}{\Gamma(1/s_n)}}|^{s_n} \end{aligned} \tag{10}$$

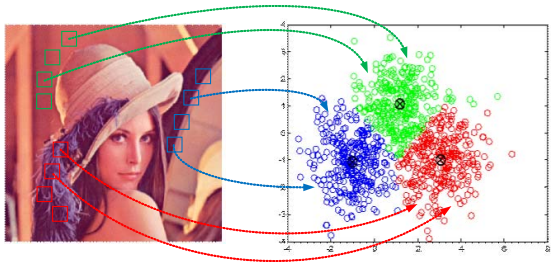


Fig. 4. Image blocks classification.

For a pixel classified into the group $\mathbf{g}_n$, (10) can be simplified as follows:

$$L_n(e) = -\log_2 \alpha_n + \log_2 \epsilon \cdot |\frac{e}{\beta_n}|^{s_n} \tag{11}$$

where β_n is a group dependent constant given by $\beta_n = \sigma_n / \sqrt{\frac{\Gamma(3/s_n)}{\Gamma(1/s_n)}}$.

As a result, in consideration of the group classification for all the pixels in a certain image region $\mathbf{R}$, the total average information cost $J(\mathbf{R})$ can be formulated by:

$$\begin{aligned} J(\mathbf{R}) &= \sum_{n=1}^{N} \sum_{p_0 \in \mathbf{g}_n} L_n(e) \\ &= \sum_{n=1}^{N} \sum_{p_0 \in \mathbf{g}_n} \left\{ -\log_2 \alpha_n + \log_2 \epsilon \cdot |\frac{e}{\beta_n}|^{s_n} \right\} \end{aligned} \tag{12}$$

Here $\mathbf{R} = \{\mathbf{g}_1 \cup \mathbf{g}_2 \cup \cdots \cup \mathbf{g}_N\}$ means all the groups in the image region $\mathbf{R}$. Through minimizing the total information cost $J(\mathbf{R})$, we can estimate the predictor coefficients a_k in (3) for a certain image region $\mathbf{R}$. Note that β_n and s_n in (11) are positive parameters, so the estimation for the predictor a_k can be carried out through solving a weighted norm optimization problem.

After minimizing the minimum rate predictor a_k, the values of the group thresholds $T_n(n = 1, 2, \cdots, N-1)$ are further optimized for a fixed set of σ_n and $s_n(n = 1, 2, \cdots, N)$. In essence, the thresholds optimization procedure is similar to the Matrix Chain Multiplication problem, and thus can be easily solved by utilizing the dynamic programing technique.

B. Adaptive Prediction Based on Block Classification

As stated before, nature images contain various kinds of textures and structural components, global predictive coding using a single predictor $a_k(i = 1, 2, \cdots, K)$ generally performs not good enough. By exploiting non-local self-similarities and structural information of image patches, we partition the whole image $\mathbf{I}$ into $S \times S$ square blocks and classify them into different M classes as depicted in Fig. 4. Then for each class, we treat all of its owned image blocks as a certain image region $\mathbf{R}^c$, and a separate minimum rate predictor $a_k^c(c = 1, 2, \cdots, M)$ is estimated correspondingly. Here c is the class index and M is the total number of classes.

Finally, for all the Cross set pixels in the whole image $\mathbf{I}$, the total information cost function $J(\mathbf{I})$ is given by:

$$J(\mathbf{I}) = \sum_{c=1}^{M} I(\mathbf{R}^c) = \sum_{c=1}^{M} \{ \sum_{n=1}^{N} \sum_{p_0 \in (\mathbf{g}_n \cup \mathbf{R}^c)} L_n(e) \} \qquad (13)$$

It should be clear that for each image class $\mathbf{R}^c$ the group thresholds $T_n (n = 1, 2, \cdots, N-1)$ are different, but all the image classes share the same groups of probability models which are governed by the parameters σ_n and s_n. So for the rest of the paper, the class dependent group thresholds are denoted by $T_n^c (n = 1, 2, \cdots, N-1)$ to avoid ambiguity.

Observing from (13), the image region classification procedure is also iteratively optimized to minimize the total average coding bits of the prediction errors in the whole image $\mathbf{I}$ simultaneously. Specifically, for each image block $\mathbf{b}_j (j = 1, 2, \cdots, B)$, a class index b_j is recorded for its belonged class. Here B is the number of image blocks and by default $b_j \in \{1, 2, \cdots, M\}$.

In comparison with the prediction coefficients a_k^c and the group thresholds $T_n^c (n = 1, 2, \cdots, N-1)$, the total amount of information bits to record the class indexes b_j is roughly large. Considering that the embedding modifications on the pixel values for reversible data hiding is generally trivial, so after embedding messages into the image pixels, we can optimize the class indexes again and obtain another b'_j. By applying entropy coding to the difference between b_j and b'_j, the extra bits to record the class indexes can be greatly reduced.

C. Embedding and Extracting

The overall Embedding and Extracting algorithms with respect to minimizing $J(\mathbf{I})$ for the Cross set pixels are described as follows. Fig. 5 illustrates the one round optimizing framework for the embedding procedure of the Cross set pixels.

Embedding

1) Divide the image $\mathbf{I}$ into $S \times S$ square blocks $\mathbf{b}_j (j = 1, 2, \cdots, B)$ and classify them into M classes. At the initialization stage, we run the K-means clustering algorithm by treating all pixels in each block as a feature vector, and the clustering results are used for the rough classification.
2) Estimate the least square predictor a_k^c for each class region $\mathbf{R}^c$ as the initial predictors without probability models classification.
3) Optimize the values of the group thresholds T_n^c in each image class region $\mathbf{R}^c$ respectively to minimize the information cost $J(\mathbf{R}^c)$ for a given predictor a_k^c and a fixed set of σ_n^2 and s_n.
4) Classify each image block $\mathbf{b}_j$ into the class $\mathbf{R}^c$ such that $J(\mathbf{b}_j)|\mathbf{b}_j \in \mathbf{R}^c$ is the smallest among all the classes. $J(\mathbf{b}_j)|\mathbf{b}_j \in \mathbf{R}^c$ means the information cost in the block $\mathbf{b}_j$ calculated by assuming that the block $\mathbf{b}_j$ belongs to class $\mathbf{R}^c$. And then update the class index by setting $b_j = c$.
5) Optimize the shape parameter s_n for each group probability model $\mathbf{g}_n$ individually by chosen one of the values from a fixed value set to minimize (13).

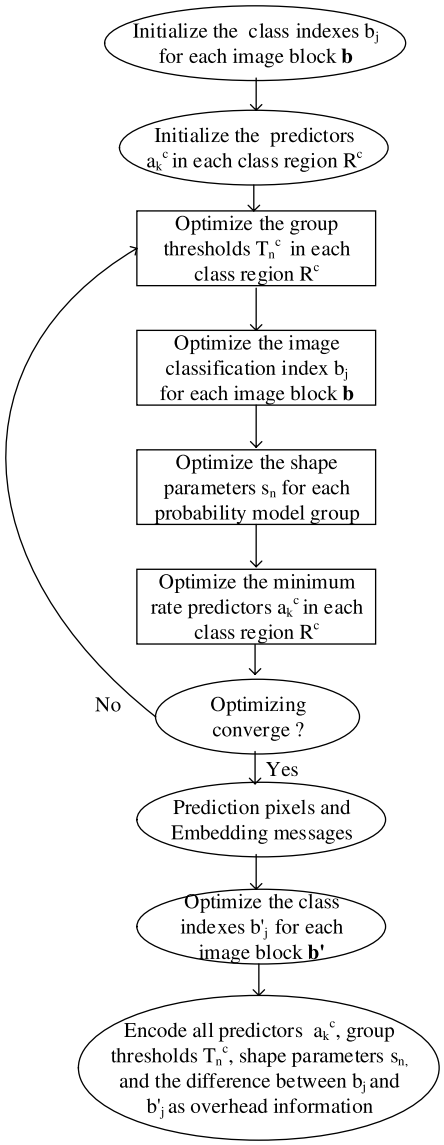


Fig. 5. One round embedding optimizing framework for Cross set pixels.

6) Optimize the minimum rate predictor a_k^c for each class $\mathbf{R}^c$ individually by minimizing (12).
7) Repeat the above procedures 3), 4), 5) and 6) until the values of the group thresholds T_n^c, the classification of the image blocks and the optimization for the minimum rate predictor converge. The maximal outer iteration number is set to 50 by default.
8) Predict the image pixels using the estimated predictors a_k^c, probability group and image block class informations accordingly. And then embedding messages into the prediction-errors e through the OHM embedding scheme. Afterwards, repeat procedure 4) once again using the modified pixel values and as a result obtain the auxiliary class indexes b'_j.
9) Encode all the predictor coefficients a_k^c, group thresholds T_n^c, shape parameters s_n as well as the difference between b_j and b'_j, and then record the coded overhead bit stream to the least significant bits (LSBs) of some reserved pixels.

Note here after the optimizing process, the prediction-error e at each pixel is rounded down to embed message:

$$e = p_0 - \lfloor \sum_{k=1}^{K} a_k^c \cdot p_k \rfloor \tag{14}$$

As for the overhead part which contains the predictor coefficients and all the other parameters, we first estimate the length of it, and then reserve enough pixels in the first couple of lines in the image to be not processed during embedding. The LSBs of these pixels are added to the message sequence to be embedded later, after embedding is done, the overhead part for the parameters are embedded into the reserved first couple of lines by simple LSB replacement.

Extracting

1) Extract the coded overhead bit stream from the LSBs of some reserved pixelsï¼Œ and then decode all the predictor coefficients a_k^c, group thresholds T_n^c, shape parameters s_n, and the difference between b_j and b'_j.
2) Repeat procedure 4) once again described in the Embedding framework to obtain the auxiliary class indexes b'_j, and together with the decoded difference between b_j and b'_j, recover the original class indexes b_j.
3) Predict the image pixels using the predictor parameters a_k^c, group and class informations respectively. And then extracting message from the marked prediction-errors e'_i through the OHM extracting scheme, meanwhile restore the original image pixel values.

After estimating the optimized predictor for the Cross set pixels, messages are embedded into the prediction errors of the Cross set pixels. Next in the second round, another optimized predictor is calculated again for the Dot set pixels while the modified Cross set pixels are used for predicting. As for the second round for the Dot set pixels, the embedding and extracting procedures are nearly the same as the first round for the Cross set pixels presented above. At the receiver side, the extraction procedures act in a reverse order, the Dot set pixels are recovered first, and the Cross set pixels are restored later in the second extraction round. This double layered embedding framework is adapted from Sachnev et al.'s work [11].

IV. Optimized Histograms Modification

A. OHM-Single and OHM-Stitched Schemes

As described in Section II-B, after generating the PE sequence by utilizing some pixel prediction strategies, the second step of PEE based reversible data hiding can be optimally conducted using the OHM method. For our minimum rate prediction strategy, although we assume that the generated PE sequence owns a mixture of probability model groups, the OHM method can be directly applied to it for embedding messages yet, if we treat the PE sequence as a single histogram regardless of the group information. We denote this direct OHM embedding scheme as OHM-Single.

Actually, the PE sequence generated by our minimum rate prediction strategy contains a group of probability distribution histograms $\{\mathbf{g}_1, \mathbf{g}_2, \cdots, \mathbf{g}_N\}$. For each group $\mathbf{g}_n (n = 1, 2, \cdots, N)$, we can run the OHM method to it and get a corresponding rate distortion embedding performance. Each group holds a different shape of histogram, which is controlled by the parameters σ_n^2 and s_n as depicted in Fig. 3. Generally, the shape of the histograms varies from a steep one to a broad one, and thus for a fixed message payload, different histogram performs differently while chosen for embedding. As a result, a resource allocation problem arises naturally, how to reasonably distribute the total message payload to these groups of histograms so that an overall minimum distortion embedding performance can be achieved?

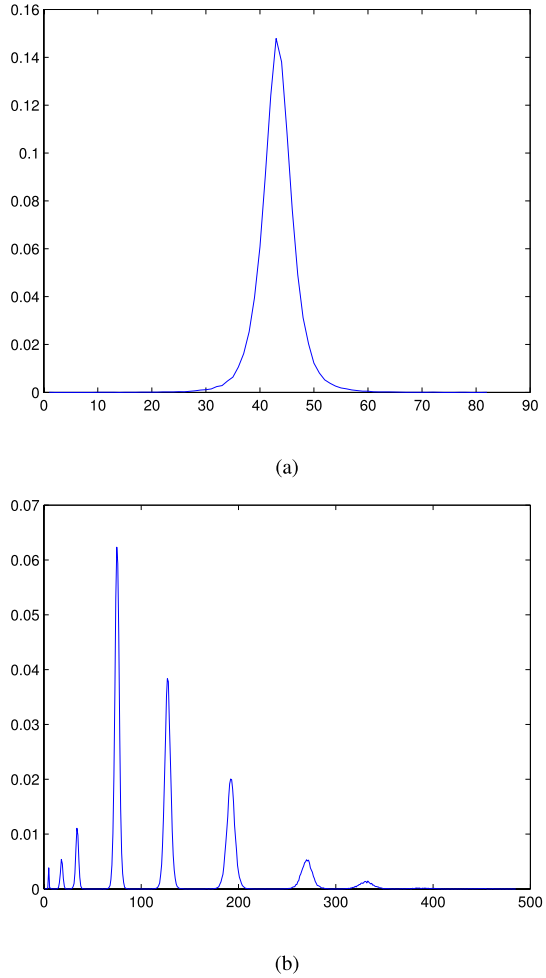


Fig. 6. Comparison of histograms of test image Lena. (a) The single histogram of Lena. (b) The stitched histogram of Lena.

Here we proposed a simple but effective solution to the resource allocation problem arose above. For each generated histogram $\mathbf{g}_n (n = 1, 2, \cdots, N)$, we assign them an offset value $o_n (n = 1, 2, \cdots, N)$. Then by translating each histogram with their corresponding offset o_n along the x-axis as depicted in Fig. 6(b), we stitch all the histograms to a big histogram. The offset values o_n is uniquely decided so that no overlap and vacant spaces occur between any two of the histograms after stitched to a big histogram.

Afterwards, the OHM embedding method can be applied to the stitched histogram, which will automatically decide the message payload distribution proportion to each sub histogram. We name this stitched histogram embedding

scheme by OHM-Stitched respectively. For the purpose of illustration, the single histogram and the stitched histogram of the Cross pixels in the test image Lena are presented in Fig. 6(a) and Fig. 6(b) accordingly.

B. Compressing Stitched Histogram for OHM-Stitched Scheme

For the OHM method described Section II-B, the frequency histogram $F_e = \{tP_e(e)\}$ of the original prediction errors e needs to be transmitted to the receiver side. Here $P_e(e)$ is the normalized probability for the prediction error and t is the total number of the prediction errors. At the sender side, we compress the histogram F_e and embed it with LSB replacements. At the receiver side, after recovering the probability distribution $P_e(e) = \{F_e/t\}$ from the decompressed histogram F_e, the recipient can estimate the optimal marginal distribution $P_Y(y)$, and then extract the message and restore the cover simultaneously.

For the OHM-Single scheme as shown in Fig. 6(a), the range of the possible prediction errors is not large, usually 0 to 90, and the shape of the histogram is unimodal. Thus to compress the frequency histogram F_e, we can first employ the differential pulse-code modulation (DPCM) technique to it, and second entropy coding will be applied to the difference values coming from the DPCM. In the experiments, the average length of the coded bits of the frequency histogram F_e for common natural images are nearly 800 to 1000.

Nevertheless, observing from Fig. 6(b), we can see that for the OHM-Stitched scheme, the stitched histogram becomes large. The max value along the x-axis of the frequency histogram F_e increased to 450, and the shape of F_e turns to a complicated multimodal form. So the conventional compressing method for the OHM-Single scheme is not suited here. Noted that for our minimum rate predictor, we expect to fit the probability distribution of the prediction error e to groups of generalized Gaussian models. Therefore, we can simulate a stitched histogram F'_e using the Gaussian model parameters σ_n and s_n and compare it to the actual generated histogram F_e of the prediction errors. By just compressing the difference between F_e and F'_e through entropy coding techniques combined with the quantization process, the length of the coded bits of the frequency histogram for the OHM-Stitched scheme can be efficiently reduced to the OHM-Single scheme's level. Fig. 7 draws the two stitched histograms for test image Lena, where we can see that they fit well to each other.

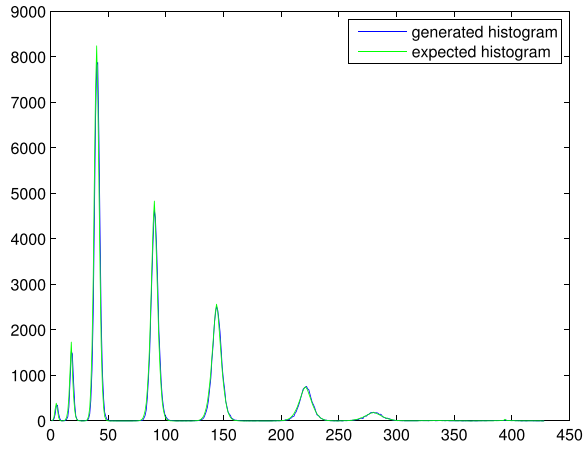


Fig. 7. Comparison between the generated stitched histogram and the expected stitched histogram of test image Lena.

V. EXPERIMENTS

A. Pixel Value Predictors Comparison

Firstly, the proposed minimum rate prediction scheme for PEE based RDH methods is compared to other six prediction schemes, namely, the median edge detector (MED) predictor [6], the simplified gradient adjusted (SGAP) predictor [24], Sachnev et al.'s rhombus predictor [11], the checkerboard based predictor (CBP) [22], the global least square (LS) predictor [25], and the local-prediction-based (LP) least square predictor [26]. TABLE I records the prediction comparison results for various test grayscale images in terms of MAE (mean absolute error), which is defined by:

$$MAE = \frac{1}{t} \times \sum_{i=1}^{t} |e_i| \tag{15}$$

where t is the total number of pixels been predicted.

For the sake of simplicity, we just predict the Cross set pixels for prediction accuracy comparisons. And for the LP prediction method [26] which predicts simultaneously with embedding, we did not embed messages while predicting. It can be seen from TABLE I that our minimum rate prediction strategy provides the best prediction accuracy among all the competitors.

Alternatively, based on the minimum rate prediction and the rate distortion formulation (1), we propose a new assessment criteria for general prediction schemes for reversible data hiding, which is called the mean histogram entropy (MHE). The MHE criteria is defined by:

$$MHE = \frac{1}{\sum_{n=1}^{N} l_n} \sum_{n=1}^{N} l_n \times H(X_n) \tag{16}$$

where l_n is the number of pixels belong to group n, and $H(X_n)$ is the entropy of the histogram of group n. N is the number of histograms. If a prediction scheme assumes a single probability model for the PE sequence, then MHE equals to the conventional entropy function $H(X)$. The prediction comparison results for various test grayscale images in terms of MHE are illustrated in TABLE II. Note here for our proposed minimum rate predictor, both the MHE values of the OHM-Single scheme and OHM-Stitched scheme are listed.

Observing from TABLE II, our minimum rate predictor provides the best prediction accuracy among all the competitors in term of the new MHE criteria, which agrees with the MAE results in TABLE I.

B. Embedding Performance Comparison

Next, to demonstrate that the MHE criteria directly indicates the embedding performance of PEE based RDH methods, we compare our proposed double-layered (the Cross set pixels

TABLE I
ONE ROUND PREDICTION ACCURACY (MAE) COMPARISONS

Image	MED [6]	SGAP [24]	Sachnev [11]	CBP [22]	LS [25]	LP [26]	Proposed
Lena	4.429	4.032	3.237	3.195	2.9616	2.645	2.521
Baboon	14.289	14.458	11.622	11.720	10.146	8.528	8.041
Barbara	9.238	8.931	7.456	6.999	5.242	2.921	2.932
Airplane	3.580	3.936	2.883	2.644	2.081	2.127	1.903
Boat	4.666	5.114	3.990	3.666	3.001	2.651	2.472
Peppers	5.359	4.829	4.051	3.963	4.104	3.449	3.287
Gold-Hill	5.453	5.698	4.531	4.526	3.995	3.811	3.573
Average	6.7166	6.714	5.396	5.245	4.504	3.733	3.533

TABLE II
ONE ROUND PREDICTION ACCURACY (MHE) COMPARISONS

Image	MED [6]	SGAP [24]	Sachnev [11]	CBP [22]	LS [25]	LP [26]	Proposed
Lena	4.555	4.420	4.115	4.100	4.008	3.893	3.841 / 3.696
Baboon	6.270	6.282	5.970	5.973	5.782	5.593	5.510 / 5.279
Barbara	5.481	5.448	5.144	4.998	4.776	4.052	4.048 / 3.828
Airplane	4.172	4.235	3.847	3.773	3.507	3.549	3.435 / 3.156
Boat	4.621	4.745	4.401	4.269	4.023	3.888	3.812 / 3.629
Peppers	4.849	4.689	4.445	4.417	4.472	4.276	4.217 / 4.084
Gold-Hill	4.885	4.951	4.619	4.610	4.436	4.421	4.339 / 4.210
Average	4.976	4.967	4.649	4.591	4.429	4.239	4.172 / 3.983

TABLE III
VALUE OF THE STANDARD DEVIATION σ_n IN THE n-TH PROBABILITY MODEL GROUP

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
σ_n	0.14	0.25	0.36	0.58	0.82	1.17	1.66	2.32	3.24	4.46	6.18	8.54	11.78	16.29	22.40	30.86

layer and the Dot set pixels layer) OHM-Single scheme and OHM-Stitched scheme to the local-prediction-based (LP) PEE scheme proposed in [26], and the classical PEE technique in [11] combined with our proposed minimum rate predictor, which is denoted by PEE-MRP. Six standard grayscale images are tested, namely Lena, Barbara, Baboon, Airplane, Boat, and Gold-Hill.

For our OHM-Single and OHM-Stitched schemes, the prediction order in Section III is set with $K = 16$, the number of groups in Section III-A and number of classes in Section III-B is set with $N = 16$ and $M = 16$. The parameters σ_n, which corresponds to the standard deviation of prediction errors in each probability model group, is fixed for all test images. TABLE III shows the values of $\sigma_n (n = 1, 2, \cdots, N)$ used in our experiments. The shape parameters s_n in each group is optimized by chosen one of the values in the following value set:

$$V = \{0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6, 1.8, 2.0, 2.2, 2.4, 2.6, 2.8, 3.0, 3.2\} \quad (17)$$

The block size for image region classification in Section III-B is set by $S = 16$.

As for the LP [26] method, the local block size is set with 12×12, the prediction order and prediction pattern is the same as ours, which is shown in Fig. 2. The LP method performs pixel predicting and message embedding simultaneously, so the OHM scheme cannot apply to it because the OHM needs the histogram of the PE sequence before embedding. The rate distortion comparison results for the four methods are depicted in Fig. 8.

Observing form Fig. 8, just provided by our minimum rate predictor, the overall gain of the PEE-MRP scheme compared with the LP method is not obvious, nearly 0.2dB to 0.3dB. And due to the extra overhead of our minimum rate predictor, the PEE-MRP scheme performs not well in low embedding rate cases. But intrinsically, the LP method embeds messages and predicts pixels one after the other, which restricts itself from using the sorting technique in Sachnev et al.'s method [11], let alone the OHM embedding scheme.

As for our OHM-Single method, an average 1.0dB PSNR gain is earned compared with the LP method. And for our OHM-Stitched method, the gains of PSNR is much higher, about 2dB to 2.5dB on average, especially under high embedding rates. Both of this indicate that to some extent, the minimum rate prediction strategy for PEE based RDH method establishes the consistency between the two steps of RDH in essence.

Recently, Gui et al. [27] proposed a high capacity adaptive embedding scheme based on generalized prediction-error expansion. They classify the cover image pixels into several levels according to their neighborhood complexity measurement, and different amount of data bits are assigned to different levels of pixels to be embedded. Smooth region pixels embed more message bits while complex texture region pixels embed less, which to some extent behaves similarly to our mixture of group probability models. We compare our OHM-Single scheme and OHM-Stitched scheme to Gui et al.'s method [27] for four standard grayscale images: Lena, Baboon, Barbara, and Airplane. Embedding

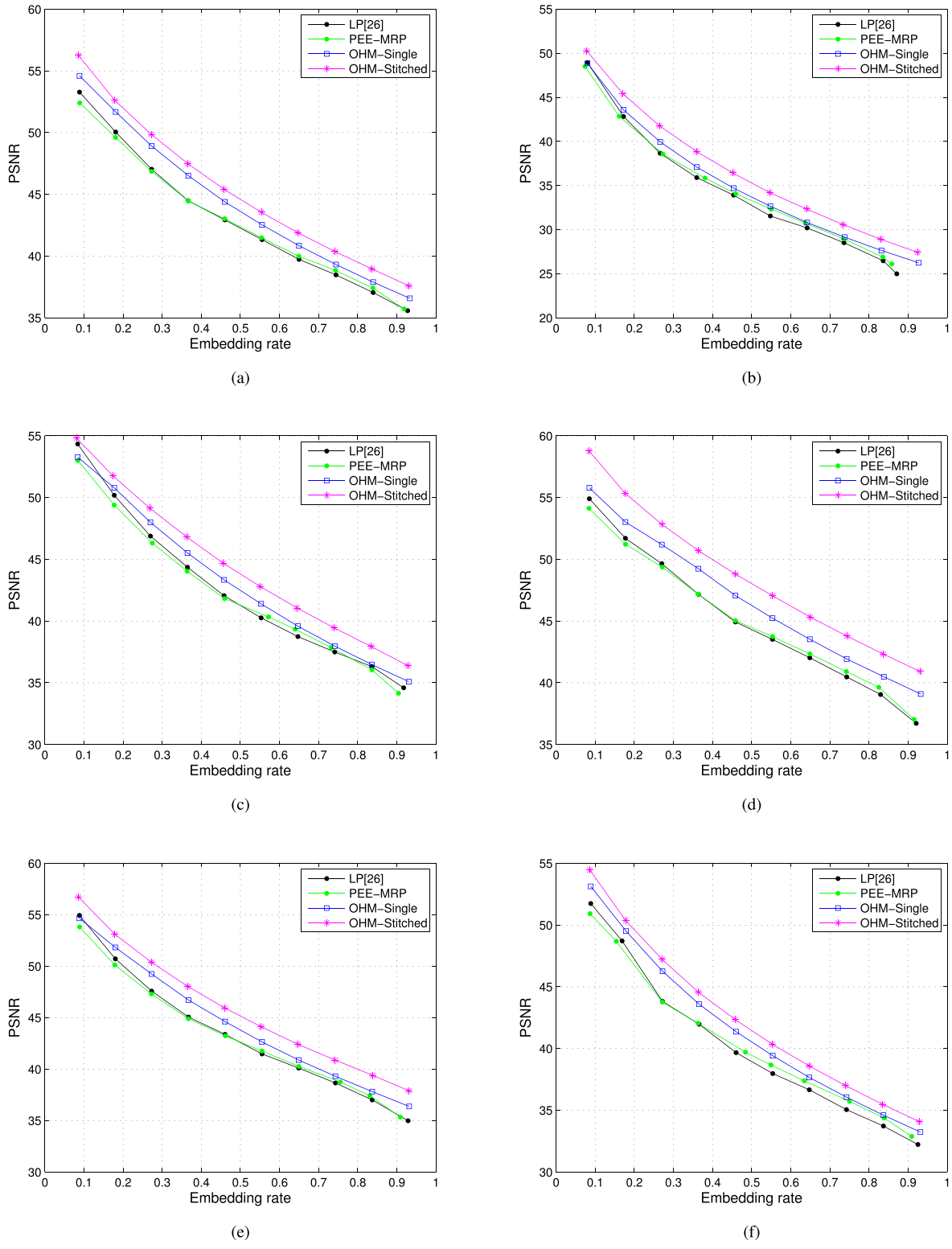


Fig. 8. Embedding performance comparisons with LP [26] and PEE-MRP. (a) Lena. (b) Baboon. (c) Barbara. (d) Airplane. (e) Boat. (f) Gold-Hill.

performance comparisons under both low embedding rates and high embedding rates are given in Fig. 9.

Seeing from Fig. 9, we note that our OHM-Single and OHM-Stitched schemes perform bad for low embedding rates in comparison with Gui et al.'s method [27]. The reason is that our proposed minimum rate prediction scheme has to record the predictor parameters a_k^c, group thresholds T_n^c, and the class index information for each image block, the total overhead length for them is nearly about 0.03 bpp to 0.04 bpp for each image. Thus under the low embedding rate scenarios, our proposed schemes' performance are overwhelmed by the extra overhead information bits. While the embedding rate

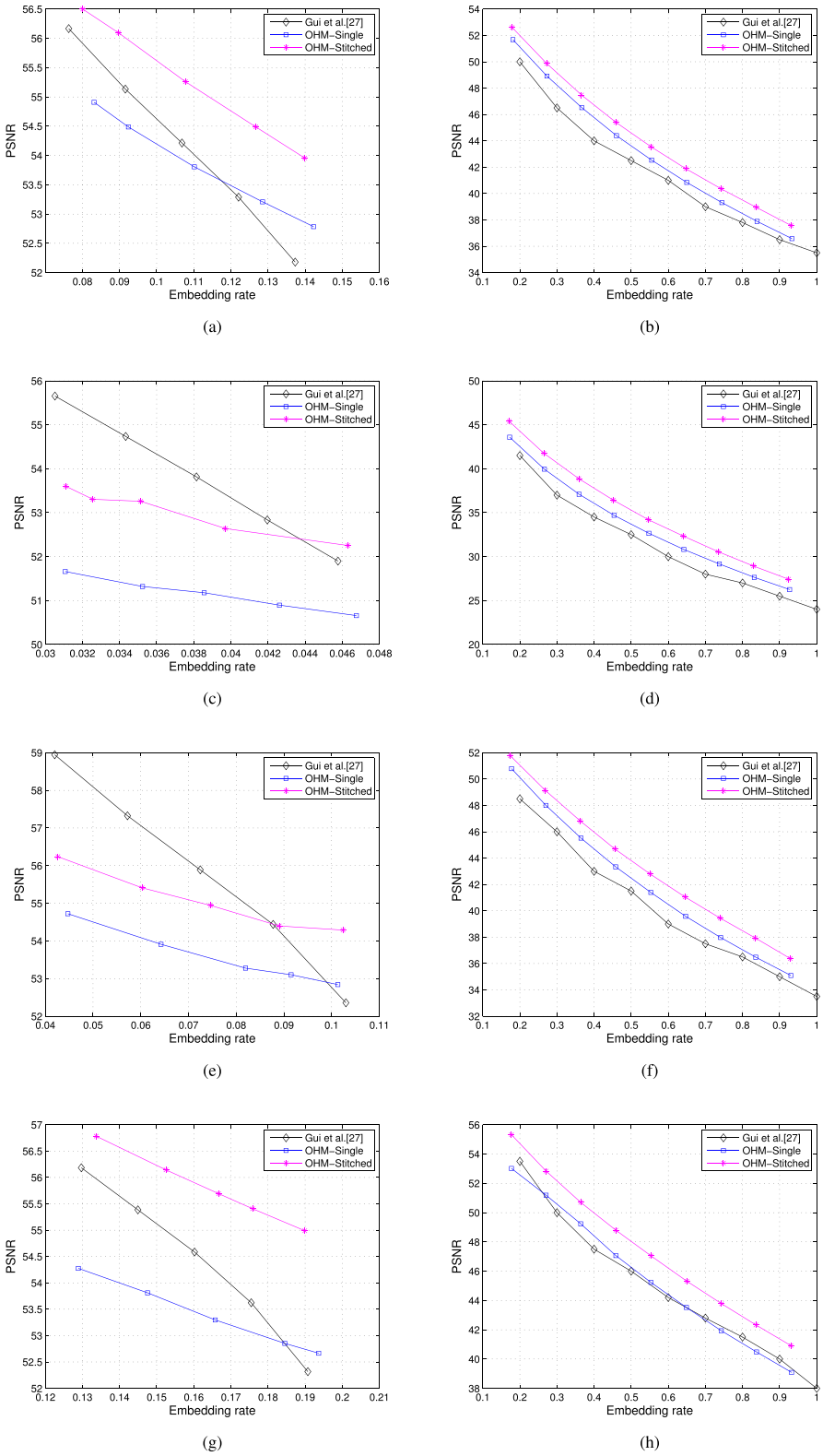


Fig. 9. Embedding performance comparisons with Gui et al. [27]. (a) Lena. (b) Lena. (c) Baboon. (d) Baboon. (e) Barbara. (f) Barbara. (g) Airplane. (h) Airplane.

grows, our OHM-Single and OHM-Stitched schemes' superior performance appears correspondingly as shown in Fig. 9.

To make it simple, for the parameters like the number of predictors K, the number of histograms N, the variances of these histograms σ_n and the corresponding shapes s_n, we just choose a fixed setting of values for all the test image. Actually, a wise and efficient scheme would be to adaptively choose or estimate these parameters for different kinds of images.

For instance, a smooth picture which contains a single texture pattern will favor less number of histograms and small variances. While for a complex picture which contains many structures and textures, more number of histograms and large variances are preferred. This subject is of much interest and value for us to be covered in our later research work.

VI. Conclusion

This paper propose a pixel prediction method for PEE based reversible data hiding schemes based on the minimum rate criterion, which establishes the consistency in essence between the two steps of PEE based reversible data hiding schemes. Previous PEE methods treat the two steps independently while they either focus on pixel prediction to obtain a sharp PE histogram, or aim at histogram modification to enhance the embedding performance for a given PE histogram. And correspondingly, a novel optimized histograms modification scheme is presented to achieve the optimal embedding performance on the generated PE sequence. Experiments demonstrate that the proposed method outperforms the previous state-of-art counterparts significantly in terms of both the prediction accuracy and the final embedding performance. Some theoretical analysis and proofs may be covered in our future work.

References

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

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

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

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

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

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

[7] M. J. Weinberger, G. Seroussi, and G. Sapiro, "The LOCO-I lossless image compression algorithm: Principles and standardization into JPEG-LS," *IEEE Trans. Image Process.*, vol. 9, no. 8, pp. 1309–1324, Aug. 2000.

[8] 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.

[9] 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.

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

[11] 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.

[12] 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.

[13] F. Peng, X. Li, and B. Yang, "Adaptive reversible data hiding scheme based on integer transform," *Signal Process.*, vol. 92, no. 1, pp. 54–62, Jan. 2012.

[14] 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.

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

[16] S.-J. Lin and W.-H. Chung, "The scalar scheme for reversible information-embedding in gray-scale signals: Capacity evaluation and code constructions," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 4, pp. 1155–1167, Aug. 2012.

[17] X. Hu, W. Zhang, X. Hu, N. Yu, X. Zhao, and F. Li, "Fast estimation of optimal marked-signal distribution for reversible data hiding," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 5, pp. 779–788, May 2013.

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

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

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

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

[22] R. M. Rad, K. Wong, and J.-M. Guo, "A unified data embedding and scrambling method," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1463–1475, Apr. 2014.

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

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

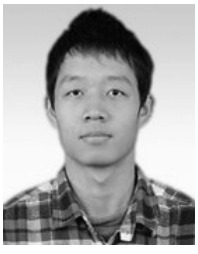
[25] S.-L. Lin, C.-F. Huang, M.-H. Liou, and C.-Y. Chen, "Improving histogram-based reversible information hiding by an optimal weight-based prediction scheme," *J. Inf. Hiding Multimedia Signal Process.*, vol. 1, no. 1, pp. 19–33, Jan. 2013.

[26] I.-C. Dragoi and D. Coltuc, "Local-prediction-based difference expansion reversible watermarking," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1779–1790, Apr. 2014.

[27] X. Gui, X. Li, and B. Yang, "A high capacity reversible data hiding scheme based on generalized prediction-error expansion and adaptive embedding," *Signal Process.*, vol. 98, pp. 370–380, May 2014.

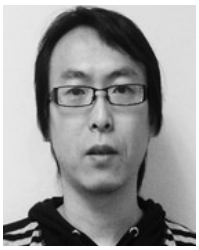
[28] I. Matsuda, N. Shirai, and S. Itoh, "Lossless coding using predictors and arithmetic code optimized for each image," in *Visual Content Processing and Representation* (Lecture Notes in Computer Science), vol. 2849. Berlin, Germany: Springer-Verlag, Sep. 2003, pp. 199–207.

[29] I. Caciula and D. Coltuc, "Improved control for low bit-rate reversible watermarking," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2014, pp. 7425–7429.



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

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



Xiaolong Li received the B.S. degree from Peking University, Beijing, China, in 1999, the M.S. degree from École Polytechnique, Palaiseau, France, in 2002, and the Ph.D. degree in mathematics from the École Normale Supérieure de Cachan, Cachan, France, in 2006. Before joining Peking University as a researcher, he worked as a postdoctoral fellow at Peking University from 2007 to 2009. His research interests are in image processing and information hiding.

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