A Lossless Data Hiding Strategy Based on Two-Dimensional Side-Match Predictions

Chi-Yao Weng[1], Sheng-Jie Wang[2], and Shiuh-Jeng Wang[2(✉)]

[1] Department of Computer Science,
National Pingtung University, Pingtung 90004, Taiwan
[2] Department of Information Management,
Central Police University, Taoyuan 33304, Taiwan
sjwang@mail.cpu.edu.tw

Abstract. The histogram-based reversible data hiding scheme (RDH) generated a one-dimensional (1D) histogram distribution. In this article, based on two-dimensional (2D) histogram distribution, a framework of reversible data hiding is proposed by using two side-match predictors, called as Forward side-match (FSM) and Backward side-match (BSM). First, by considering each predicted pixel value, we use two side-match predictors to obtain two prediction error distributions. A slope meter is computed by the differencing of two distributions. Then, a two dimensional histogram is generated by composing of BSM distribution and slope meter. Based on the 2D, more specified spaces can be found to enhance the performance. The experimental results demonstrated that our proposed scheme can achieve better performance in terms of both marked image quality and embedding capacity than that of conventional works.

Keywords: Reversible data hiding (RDH) scheme · Two dimensional histogram distribution · Side-match predictor

1 Introduction

Data hiding is an advanced technology for protecting the message with integrity and security from data transmission in a public network [1]. The secret information is capable of being hidden into some media, such as an image, audio or video data, using the well-known data hiding scheme. In general, for different target, the data hiding scheme can be classified into two categories. One is non-reversible data hiding scheme [2–4]. The non-reversible data hiding scheme aims to protect the hidden data for secret communication. The conventional LSBs (least signification bit substitution) is a simple and easy technology of non-reversible data hiding scheme. For example of 1-bit LSBs method, a secret bit is to be embedded into 1-rightmost bit of the cover pixel. The other is reversible data hiding (RDH) scheme [5–16]. The goal of RDH algorithm is to obtain both cover media and hidden data that can be recovered and extracted from stego media. It indicates that the distortion of cover media is intolerable when secret message is extracted. RDH can be widely applied into some applications, such as military, medical, or legal document.

Many RDH techniques have been developed in present. It can be divided into three categories, namely, histogram-based [5–8], difference expansion [9–12], and prediction

error [13–16]. Ni et al. first proposed a reversible data hiding based on histogram shifting in 2006 [5]. In their method, the histogram distribution is fist generated from an input image. After that, shift and conceal the secret bits into the selected range from histogram distribution. Ni's approach performed a good quality in the stego image because pixel shifting and data concealing are only shifted by one unit. Although histogram-based approach is simplicity and provides high image quality, the number of embedding capacity is limited in the height of peak point. Later on, more efficient algorithms have been invented for improving the embedding capacity limitation [6–8].

The second category is difference expansion. In 2003, Tian proposed a reversible data hiding scheme based difference expansion (DE) [9]. Tain's method expands the two pixel differencing for data concealing. The maximal embedding ratio based on DE is near to 0.5 bpp (bit per pixel) for one layer embedding. Some researchers are focused on investigation and developing high-fidelity reversible data hiding scheme [10–12].

The histogram-based approach and difference expansion method can perform lossless data hiding, but their performance might be limited within the media context. A basic idea is to review their drawback and develop suitable strategy for promoting the performance in terms of embedding capacity and image quality. The prediction error is a good strategy for both histogram-based approach and difference expansion method to devise a new method of improving the performance. A local prediction reversible data hiding combined prediction error and difference expansion is proposed by Drago and Coltur in 2014 [13]. Drago and Colutr's approach utilized least square predictor to obtain the prediction error, then, expend the prediction error for data hiding. This approach has better performance than that of DE method. Later on, more researchers are developed the efficient approaches for obtaining higher performance [14–16].

In reviewing above-mention reversible data hiding scheme, a framework of reversible data hiding is proposed. Our main contribution is to build 2D histogram distribution based on two side-match predictors and slope meter. 2D histogram distribution can collect more information and find out more spaces for date embedding, resulting in high embedding capacity. Our proposed can be extended into multi-dimensional framework in future.

The rest of the paper is organized as follows. The related works about histogram shifting scheme are briefly introduced in Sect. 2. Section 3 presents our proposed scheme in detail. The experimental result is illustrated in Sect. 4. Finally, the conclusion is concludes in Sect. 5.

2   Related Works

In this subsection, we will review the reversible data hiding scheme based on histogram shifting. The histogram shifting scheme, named as histogram-based, is introduced in 2006. Ni et al.'s approach found out that each cover image has its pixel distribution. Firstly, they applied the statistic function to account each pixel number, the histogram distribution is thus generated. Next, find out the range of peak point and zero point

from this distribution. In the end, shift and conceal the data into the range of peak point and zero point. The detail algorithm of histogram-based is shown as below.

Input: Cover image *CI*, each pixel value $x \in [0.255]$, secret message $SM = b_1b_2b_3 \ldots b_n$, $b_i \in \{0,1\}$.
Output: Stego image *SI*, a pair information (peak point, zero point).

- Step 1: Generate a histogram distribution $H(x)$ from an inputted image. Where, Fig. 1 is an example of histogram distribution using Lena image.
- Step 2: Find out a pair-data (P, Z) from the histogram distribution. Notable, P indicates the maximum number of pixels in this distribution. Z means no pixel or minimum pixel in this distribution.
- Step 3: Shift all the pixels x within the range of peak point and zero point by one unit. Note, the Fig. 2 shows the result after pixel shifting operation. The shifting function is given as following.

$$\begin{cases} x' = x + 1, \text{if } Z > P \text{ and } x \in [P + 1, Z] \\ x' = x - 1, \text{if } P > Z \text{ and } x \in [Z, P - 1] \end{cases} \quad (1)$$

- Step 4: Scan all pixels in the cover image, once the pixel value meets the P value, and then, check the secret bits. If the secret bit is "0", the pixel value remains unchanged. If the secret bit is "1", the pixels value is changed by one unit according to following data concealing function

$$\begin{cases} x' = x + 1, \text{if } Z > P \\ x' = x - 1, \text{if } P > Z \end{cases} \quad (2)$$

Finally, output all the scanned pixel as the stego pixel (stego image). Notably, the stego pixel can be performed the processes of data extracting and pixel recovering since a pair-data (P, Z) is received. For the data extracting, the secret bit will be extracted when the inputted pixel meets the value of P or $P + 1$. In other words, the hidden bits can be successfully retrieved. For the concept of pixel recovering, each inputted pixel can run the inverse of shifting operation to recover the pixel value as an original pixel.
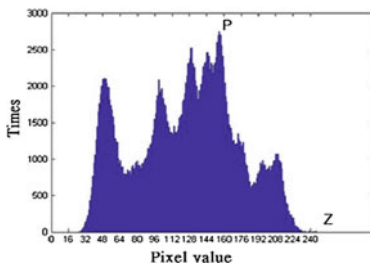

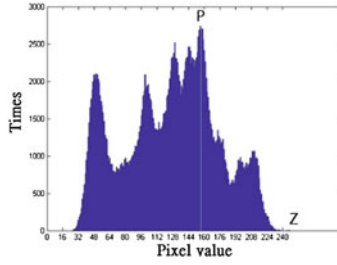**Fig. 1.** The example of pixel distribution using Lena as cover image

Fig. 2. The example of pixel shifting using Fig. 1.

3   Our Proposed Method

In this section, we introduce our reversible data hiding scheme based on 2D histogram modification in detail. The 1D histogram-based approach applied one factor, such as pixel value or prediction error, to generate a factor histogram distribution. Then, modify the highest value in the factor distribution for data concealing. For receiving more efficiency, our proposed scheme uses two prediction error to produce a 2D histogram distribution. More specified cases are to be considered for pursuing more spaces for promoting the embedding capacity. The detail of our data embedding algorithm is shown as below.

Input: Cover image CI with size $M \times N$, $CI = I_{i,j}$, $i = 0$ to M-1, $j =$ to N-1, secret message $SMsg = b_1 b_2 b_3 \ldots b_n$, $b_i \in \{0,1\}$.

Output: Marked-image MI, two pair information (PP_0, PZ_0) and (NP_0, NZ_0).

- Step 1: Scan the cover image and run two side-match predictor methods, Forward side-match (FSM) and Backward side-match (BSM), to predict the pixel vale in the image.

  - Step 1.1: The FSM prediction algorithm is shown the subsection, and then, compute the prediction error PE_FSM$_{i,j}$ = $PI_{i,j}$ - $I_{i,j}$.
    - Step 1.1.a: If $i = 0$ and $j = 0$, then $PI_{i,j} = I_{i,j}$ -128.
    - Step 1.1.b: If $i = 0$, then $PI_{i,j} = I_{i,j} - I_{i,j-1}$.
    - Step 1.1.c: If $j = 0$, then $PI_{i,j} = I_{i,j} - I_{i-1,j}$.
    - Step 1.1.d: Else, $PI_{i,j} = I_{i,j}$ - int (average ($I_{i-1,j}$ +$I_{i-1,j-1}$ + $I_{i,j-1}$)).

  - Step 1.2: The Backward side-match (BSM) prediction algorithm is shown the subsection, and calculate the prediction error PE_BSM$_{i,j}$ = $PI_{i,j}$ - $I_{i,j}$.
    - Step 1.2.a: If $i = 0$ and $j = M$-1, then $PI_{i,j} = I_{i,j}$ -128.
    - Step 1.2.b: If $i = 0$, then $PI_{i,j} = I_{i,j} - I_{i,j+1}$.
    - Step 1.2.c: If $j = 0$ and $j \neq 0$, then $PI_{i,j} = I_{i,j} - I_{i-1,j}$.
    - Step 1.2.d: Else, $PI_{i,j} = I_{i,j}$ - int (average ($I_{i-1,j} + I_{i,j-1}$)).

  - Step 1.3: Obtain each slope meter SM from computing the difference of PE_BSM$_{i,j}$ and PE_FSM$_{i,j}$, where $SM_{i,j}$ = PE_FSM$_{i,j}$ — PE_BSM$_{i,j}$.

Step 2: Let $(SM_{i,j}, PE\_FSM_{i,j})$ denotes a 2D value, and collect all 2D values to generate a 2D histogram distribution, $H(SM_{i,j}, PE\_FSM_{i,j})$.

Step 3: Find two pair information (PP_0, PZ_0) and (NP_0, NZ_0) from 2D histogram distribution. Where a pair information (PP_0, PZ_0) is selected in the positive regions of 2D histogram distribution. In the same way, a pair information (NP_0, NZ_0) is picked by the negative positive regions of 2D histogram distribution.

Step 4: Shift all the prediction error $PE\_FSM_{i,j}$ by 1 unit according to following function.

$$PE\_FSM'_{i,j} = \begin{cases} PE\_FSM_{i,j} + 1, PE\_FSM_{i,j} \in [PP_0 + 1, PZ_0] \\ PE\_FSM_{i,j} - 1, \ \ PE\_FSM_{i,j} \in [NZ_0, NP_0 - 1] \end{cases} \quad (3)$$

Step 5: Scan all the $PE\_FSM_{i,j}$ and fetch each secret bits from secret message $SMsg$, and then, conceal the secret bitstring into $PE\_FSM_{i,j}$ value while $SM_{i,j}$ precisely equals to "0" and the $PE\_FSM_{i,j}$ equals to the peak values, PP_0 and NP_0. The data concealing function is listing as below.

$$PE\_FSM'_{i,j} = \begin{cases} PE\_FSM_{i,j} + 1, PE\_FSM_{i,j} = PP_0 \text{ and } b_i = 1 \\ PE\_FSM_{i,j}, PE\_FSM_{i,j} = PP_0 \text{ and } b_i = 0 \\ PE\_FSM_{i,j}, PE\_FSM_{i,j} = ZP_0 \text{ and } b_i = 0 \\ PE\_FSM_{i,j} - 1, PE\_FSM_{i,j} = ZP_0 \text{ and } b_i = 1 \end{cases}$$

$$PE\_FSM'_{i,j} = \begin{cases} PE\_FSM_{i,j} + 1, PE\_FSM_{i,j} = PP_0 \text{ and } b_i = 1 \\ PE\_FSM_{i,j}, PE\_FSM_{i,j} = PP_0 \text{ and } b_i = 0 \\ PE\_FSM_{i,j}, PE\_FSM_{i,j} = ZP_0 \text{ and } b_i = 0 \\ PE\_FSM_{i,j} - 1, PE\_FSM_{i,j} = ZP_0 \text{ and } b_i = 1 \end{cases} \quad (4)$$

Step 6: Recover the marked prediction error value $(PE\_FSM_{i,j})$ to be a stego-pixel (SPI) according the inverse of Forward side-match (FSM) predictor, where the inverse algorithm is given as below.

Step 6.1: If $i = 0$ and $j = 0$, then $SPI_{i,j} = PE\_FSM_{i,j} + 128$.

Step 6.2: If $i = 0$, then $SPI_{i,j} = PE\_FSM_{i,j} + I_{i,j-1}$.

Step 6.3: If $j = 0$, then $SPI_{i,j} = PE\_FSM_{i,j} + I_{i-1,j}$.

Step 6.4: Else, $SPI_{i,j} = PE\_FSM_{i,j} + int (average (I_{i-1,j} + I_{i-1,j-1} + I_{i,j-1}))$.

All the marked prediction error values are fully to be recovered, the marked image (MI) is therefore outputted.

The data extracting procedure is similar with the data embedding procedure. Two prediction methods, FSM and BSM, are also used in the extracting process. The extracting process is the inverse of the data embedding procedure. After running the data extracting procedure, the secret message is completely obtained and the image is successfully recovered.

4 Experiment Results

In section, we demonstrate the simulations of our proposed scheme. Five 512×512 images, including Lena, Airplane, Pepper, Boat, and Goldhill (See Fig. 3) are used in our simulations. The secret bitstring in our simulation is generated by the random number generator. We also employ the commonly measure function peak-signal-to-noise-ratio (PSNR) and MSE as a evaluate tool for comparing the visual quality between the cover image and marked image. The measure functions are given as following.

$$\text{PSNR(db)} = 10 \times log_{10}\left(\frac{Max^2}{MSE}\right) \tag{5}$$

$$\text{MSE} = \frac{1}{W \times H}\sum\nolimits_{x=1}^{W}\sum\nolimits_{y=1}^{H}\left((I(x,y)) - (I'(x,y))\right)^2 \tag{6}$$

where W and H are defined as the width and height of the image. I(x, y) and I'(x, y) indicates the values of cover image and marked image, respectively. The Max means the maximum value of cover image. Here, our cover image is gray level image, thus, the Max value in gray level is $255(= 2^8)$.
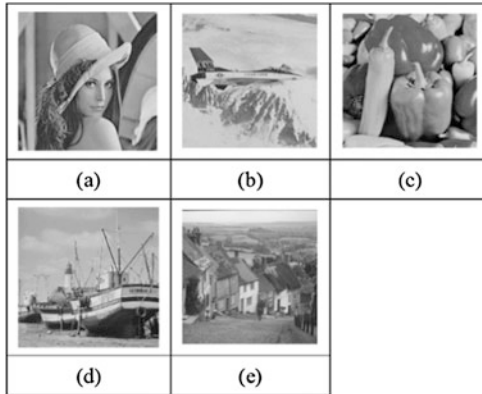

Fig. 3. The five cover images. (a)Lean; (b)Airplane; (c)Pepper; (d)Boat; (e)Goldhill

To obtain a better knowledge of how different cover image impact the performance of the proposed RDH scheme, we compare our experimental results with the recent RDH scheme, the comparison result is illustrated in Table 1. From this comparison, our scheme has better performance in terms of high embedding capacity and good image quality than that of existing RDH methods. The side match predictor applied the neighbor pixels to predict the prediction value. Generally speaking, in a nature image, the current pixel is very similar with its neighbor pixels. Assume that we use side match

manner to predict the value, we can obtain an accuracy prediction. It leads that more and more prediction errors are falling into zero space. This is a reason that our side match predictor explores the advantage of prediction error to create more space for data concealing.

Table 1. The performance comparison with various approaches

	Ni et al.'s [5]	Tsai et al.'s [7]	Li et al.'s [14]	Our approach
Lean	0.02	0.18	0.23	0.25
Airplane	0.07	0.25	0.30	0.29
Pepper	0.02	0.18	0.24	0.23
Boat	0.04	0.17	0.19	0.22
Goldhill	0.02	0.13	0.19	0.17
Avag. Capacity (bpp)	0.034	0.182	0.230	0.232
Avag. PSNR (db)	48.30	48.7+	48.47	49.32

5   Conclusion

In this paper, based on two side match predictors and two peaks histogram embedding, an efficient reversible data hiding scheme is proposed. The key point is to build 2D histogram based on two side match predictors and slope meter. The slope meter is used to estimate the differencing of two prediction errors. Based on accurate prediction and selecting two peak points, our proposed scheme provides more spaces for enhancing embedding capacity, the evidence can be seen in the performance evaluation. From the experimental result, it shows that our proposed approach is superior over some histogram-based works.

References

1. Bender, W., Gruhl, D., Morimote, N., Lu, A.: Techniques for data hiding. IBM Syst. J. **35** (3–4), 313–316 (1996)
2. Chan, C.K., Chen, L.M.: Hiding data in images by simple LSB substitution. Pattern Recogn. **37**(3), 469–474 (2004)
3. Sun, H.M., Weng, C.Y., Wang, S.J., Yang, C.H.: Data embedding in image-media using weight-function on modulo operation. ACM Trans. Embed. Comput. Syst. **12**(2), 1–12 (2013)
4. Maleki, N., Jalail, M., Jahan, M.V.: Adaptive and non-adaptive data hiding methods for grayscale images based on modulus function. Egypt. Inform. J. **15**(2), 115–127 (2014)
5. Ni, Z.C., Shi, Y.Q., Ansari, N., Su, W.: Reversible data hiding. IEEE Trans. Circuits Syst Video Technol. **16**(3), 354–362 (2006)

6. Tai, W.L., Yeh, C.M., Chang, C.C.: Reversible data hiding based on histogram modification of pixel differences. IEEE Trans. Circuits Syst. Video Technol. **19**(6), 906–910 (2009)
7. Tsai, P., Hu, Y.C., Yeh, H.L.: Reversible image hiding scheme using predictive coding and histogram shifting. Sig. Process. **89**(6), 1129–1143 (2009)
8. Tain, J.: Reversible data embedding using a difference expansion. IEEE Trans. Circuits Syst. Video Technol. **13**(8), 890–896 (2003)
9. Kim, H.J., Sachnev, V., Shi, Y.Q., Nam, J., Choo, H.G.: A novel difference expansion transform for reversible data embedding. IEEE Trans. Inf. Forensics Secur. **3**(3), 456–465 (2008)
10. Alattar, M.: Reversible watermark using the difference expansion of a generalized integer transform. IEEE Trans. Image Process. **13**(8), 1147–1156 (2004)
11. Sachnev, V., Kim, H.J., Nam, J., Suresh, S., Shi, Y.Q.: Reversible watermarking algorithm using sorting and prediction. IEEE Trans. Circuits Syst. Video Technol. **19**(7), 989–999 (2009)
12. Wu, H.C., Lee, C.C., Tsai, C.S., Chu, Y.P., Chen, H.R.: A high capacity reversible data hiding scheme with edge prediction and difference expansion. J. Syst. Softw. **82**(12), 1966–1973 (2009)
13. Dragoi, I., Coltuc, D.: Local-prediction-based difference expansion reversible watermarking. IEEE Trans. on Image Processing **23**(4), 1779–1790 (2014)
14. Li, X., Yang, B., Zeng, T.: Efficient reversible watermarking based on adaptive prediction-error expansion and pixel selection. IEEE Trans. Image Process. **20**(12), 3524–3533 (2011)
15. Yang, C.H., Tsai, M.H.: Improving histogram-based reversible data hiding by interleaving predictions. IET Image Process. **4**(4), 223–234 (2010)
16. Wang, S.Y., Li, C.Y., Kuo, W.C.: Reversible data hiding based on two-dimensional prediction errors. IET Image Process. **7**(9), 805–816 (2013)