Journal of Information Hiding and Multimedia Signal Processing ©2014 ISSN 2073-4212
Ubiquitous International Volume **5**, Number **3**, July **2014**

Reversible Data Hiding Techniques Using Multiple Scanning Difference Value Histogram Modification

John Marin and Frank Y. Shih

Department of Computer Science
New Jersey Institute of Technology
GITC Building, Room 4400 University Heights, Newark, NJ 07102-1982, USA
shih@njit.edu

Received October, 2013; revised January, 2014

Abstract. *We present an improved histogram modification reversible data hiding algorithm using multiple scanning techniques. Using the size of the desired payload, an embedding order based on the embedding level and scanning technique is determined. Difference value histogram modification is then repeatedly performed on the image using the embedding order, and the overhead information is stored within the image. Experimental results indicate that our algorithm provides on average a 1.12 bpp payload capacity with 30dB PSNR and a 1.92 bpp maximum payload capacity. This is a 3% and 120% improvement, respectively, over using a single scanning technique.*
Keywords: Reversible data hiding, difference value, histogram modification

1. **Introduction.** Data hiding is the process of hiding data within some media. In digital image media this means altering the values of the original image to embed additional information. While there are many straightforward implementations of data hiding, such as least significant bit based hiding; these methods do not allow for the recovery of the original image [1]. Reversible data hiding involves embedding data into the image in a manner that allows for the retrieval of the embedded data as well as the exact original image. This aspect is desirable for fields that require high precision of an image.

Many reversible data hiding schemes have been proposed [2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12]. Tian [2] proposed a difference expansion data hiding scheme, where the difference and average values of two neighboring pixels are calculated and the secret data to be embedded are appended to a difference value represented as a binary number. Alattar [3], Kim et al. [4] and Weng [5] further extended Tian's work. Ni et al. [6] utilized the histogram of the pixels in the cover image to design a reversible hiding scheme, where the pixels between the peak and zero pair are modified in the embedding process and the pixel in the peak point is used to carry a bit of the secret message.

We present an improved histogram modification reversible data hiding algorithm using multiple scanning techniques. Two performance metrics are important to consider with a reversible data hiding technique. The first is the maximum payload, which is the maximum amount of data that can be embedded into an image. The second is visual quality, which is measured using the PSNR (Peak Signal to Noise Ratio) between the original image and the embedded image.

This paper is organized as follows. In Section 2, we provide an overview of the proposed multiple scanning histogram modification algorithm. The single scanning histogram modification algorithm and multiple scanning histogram modification algorithms are described in Sections 3 and 4, respectively. Section 5 presents an algorithm for determining the embedding level and scan order. Section 6 shows experimental results. Conclusions are drawn in Section 7.

2. **Overview of the Multiple Scanning Histogram Modification Algorithm.** Our proposed algorithm consists of two stages: encoding the payload within an image to produce the cover image and decoding the cover image to retrieve the payload and the original image. We use three different scanning techniques for the construction of the difference histograms in order to maximize the payload capacity. The scanning techniques that we use are a horizontal, vertical, and diagonal scanning technique. The horizontal scanning technique is the same as that used in Zhao et al.'s algorithm [11]. Fig. 1 shows the scanning order on a 4 × 4 image. The scanning order uses adjacent pixels since these are more likely to have similar values and as a result will result in a difference histogram that has most of its values near 0.

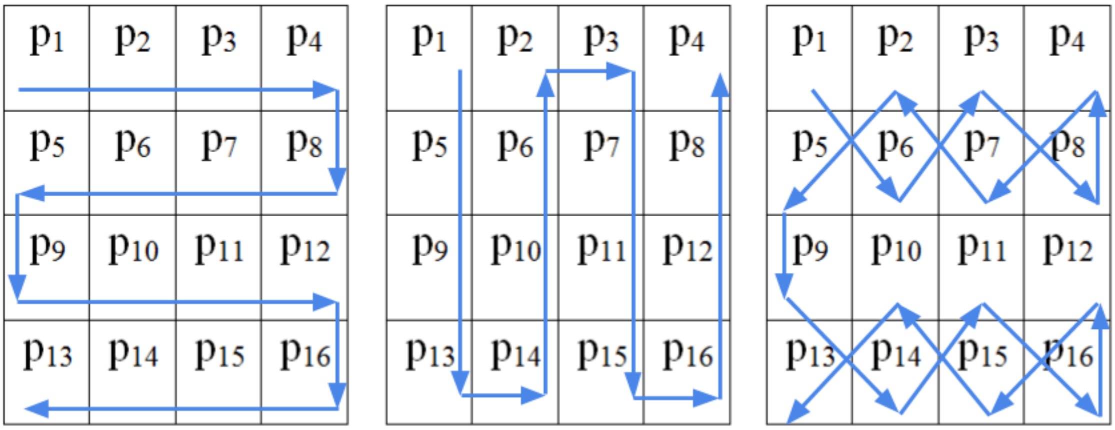

FIGURE 1. Pixel scan order for a 4 × 4 image

Let us introduce the following symbols and notations that will be used throughout this paper.
I: The original image
L: The layer of embedding, i.e. the number of times the histogram shifting algorithm has been applied to the image
IL: An image that has been embedded L times
P: The payload embedded in the image
LM: Location map of pixels that can be used in data embedding
LMC: The compressed location map

2.1. **The Embedding Process of our Reversible Data Hiding Algorithm.**

1. Determine histogram shift order based on the payload.
2. Pairs of values corresponding to the scanning method and EL are outputted (discussed in detail in Section 5). An integer parameter called embedding level EL (EL 0) is involved to control the hiding capacity. A larger EL reflects that more secret data can be embedded.
3. Calculate the minimum number of pixels required to embed LMc .
4. Segment the image into I1 and I2 .

5. Calculate LM of I1 .
6. Extract the least significant bits (LSB) of I2 and concatenate them with the payload.
7. Compress LM using arithmetic coding and embed it within the LSB of the pixels of I2.
8. Use the given scanning method to calculate the difference histogram of I1 minus the bits in LM. Concatenate the key, I2, and P to make PM .
9. Shift the difference histogram for the given EL and embed PM.
10. Apply the new difference histogram to the image.
11. Repeat steps 3-10 until all of the histogram shifts are applied and the payload is embedded. Save the scanning method, EL, and the key of the last histogram shift as external data.

2.2. The Extracting Process of our Reversible Data Hiding Algorithm.

1. Use the scanning method, EL, and key provided.
2. Extract LMc from IL and decode into LM.
3. Calculate the embedded difference histogram of IL.
4. Recover the original difference histogram and extract PM.
5. Separate PM into the LSB of I2, the next decoding method, EL, and key (if applicable), and P.
6. Use the original difference histogram to calculate IL-1.
7. If PM contains a decoding method, EL, and key, use these values and go to step 2.
8. Concatenate all of the bits of P.

3. **Single Scanning Histogram Modification.** Zhao et al. [11] proposed a reversible data hiding algorithm using a single scanning technique to produce the difference histogram. The scanning method used is the inverse S, which is the same as the horizontal scanning order in our algorithm. Since one scanning technique is used, a single difference histogram is produced. Fig. 2 shows the Lena image and its difference value histogram showing difference values between -20 and 20 using inverse S scanning.

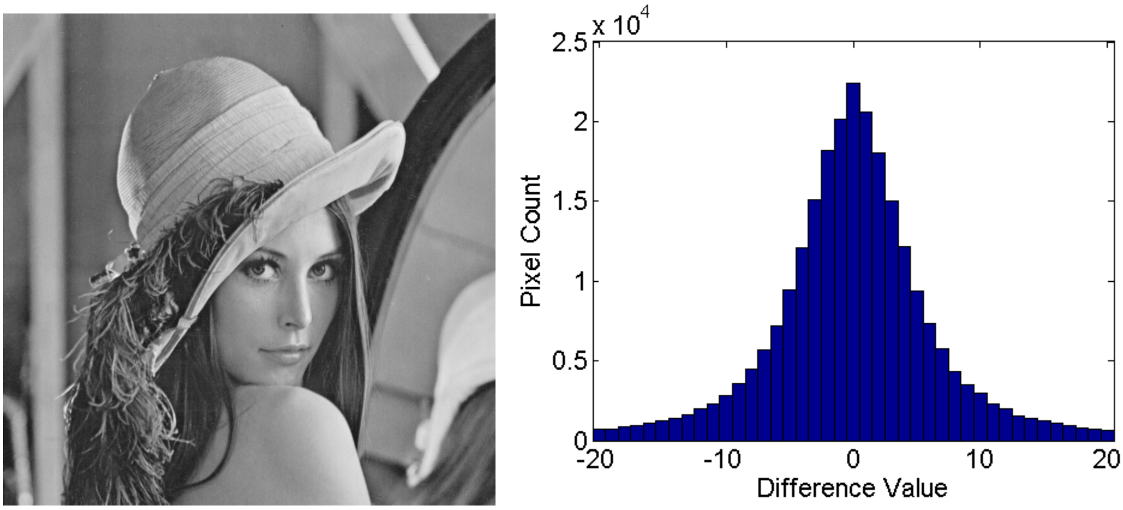

Figure 2. Lena image and the difference value histogram using horizontal scanning

The difference histogram of the Lena image clearly shows that the most differences are around 0 with values farther from 0 becoming less frequent. Using this characteristic, Zhao et al. uses pixels with difference values close to 0 for data embedding. Data is embedded by shifting the difference histogram, such that two spaces are available for each difference value that has data embedded. This is to accommodate the binary values 0 and 1. Fig. 3 shows the shifted Lena difference value histogram and after data is embedded at each embedding level.

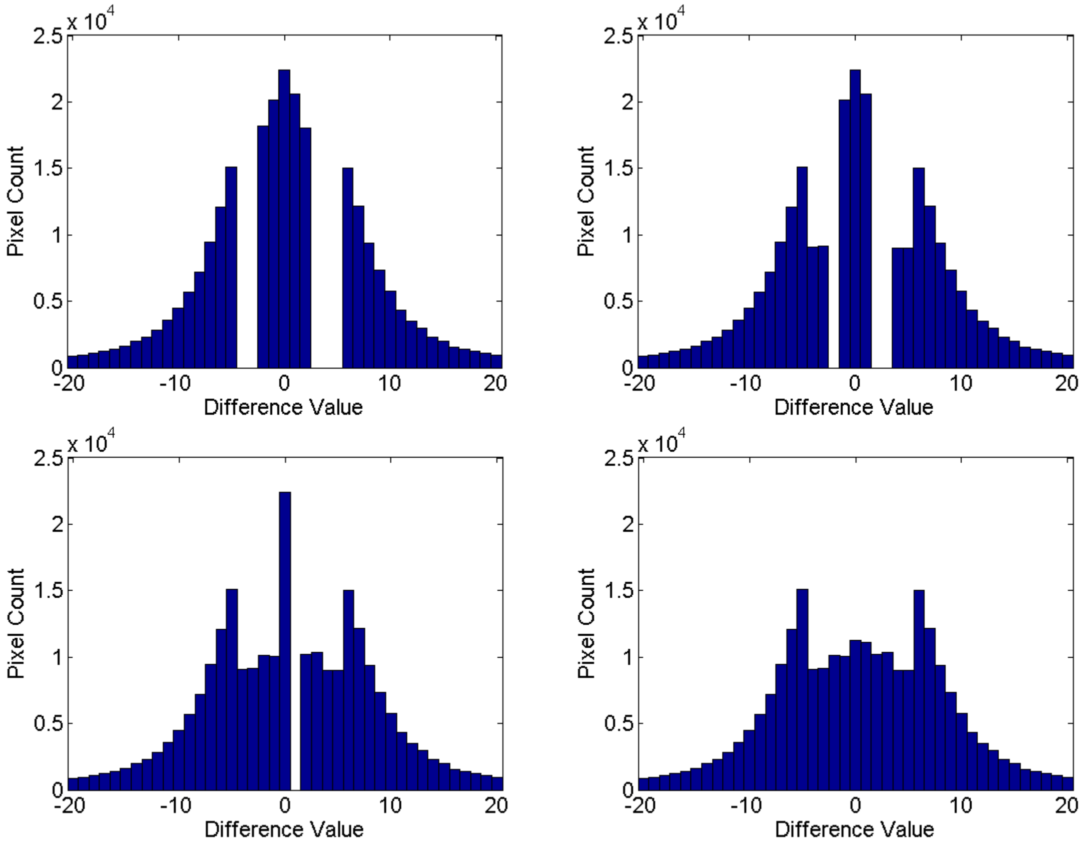

FIGURE 3. (a) Shifted Lena difference histogram and embedded histogram at (b) EL=2, (c) EL=1, and (d) EL=0

The maximum capacity and PSNR of Zhao et al.'s algorithm depends on the embedding level and consequently the number of histogram shifts that are made to the difference values. Since the EL can be increased such that all pixels are used for data embedding, the theoretical maximum of an images data embedding capacity is 1bpp. However, this maximum cannot be achieved using this technique because overhead information is required and some pixels may not be embeddable.

Since the data embedding process can result in the overflow (values ¿ 255) or underflow (values ¡ 0), the pixels whose values can take on these values after embedding cannot be altered. The pixels that are vulnerable to overflow or underflow are those in the intervals [0, EL] and [255 EL, 255]. Not only must these pixels remain unchanged, but they must have their locations to be saved and relayed to the decoder, so they are skipped during the decoding process as well. Zhao et al. used an M × N binary location map LM to prevent overflow and underflow. Before data is embedded, all pixels that are outside the valid range are excluded and have a value of 1 recorded in the LM; otherwise, a 0 is recorded. The LM is then compressed using arithmetic coding. Since the LM must be known to the decoder before the difference values are calculated, the compressed LM must be stored separately in the image. The image is split into two parts I1 and I2, where I1 is used to embed the payload and I2 is used to store LMC. LMC is stored in I2 by replacing the LSBs of I2 with LMC. The LSBs of I2 are then concatenated with the payload and the difference histogram shifting algorithm is performed on the pixels of I1.

The maximum data embedding capacity of Zhao et al.'s algorithm is limited by two factors, the number of pixels available at each embedding level and the amount of data overhead caused by potential overflow and underflow. With an increasing EL, the number of pixels that are ineligible for embedding increases as well as the size of LMC which reduces the capacity for payload bits at each embedding level. In addition, the difference value histograms at higher embedding levels have shorter peaks. These two factors result in a peak embedding capacity for images as shown for the Lena image in Fig. 4.

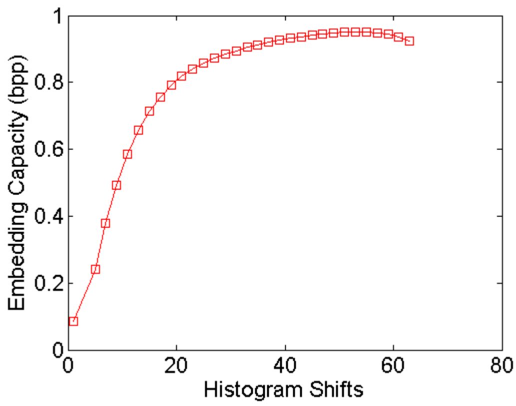

FIGURE 4. Embedding capacity of the Lena image using Zhao et al.'s algorithm

Fig. 4 indicates the embedding capacity using the single scanning method increases rapidly with a small number of histogram shifts and has greatly diminishing returns for higher histogram shifts. The embedding capacity gradually levels off and declines, which is when the overhead information is greater than the improvement in payload capacity. As each pixel can only hold 1 bit of payload data, Zhao et al.'s algorithm has a hard upper bound of 1bpp. This limitation inspired our proposed technique which attempts to break this capacity ceiling while meeting or exceeding the cover image quality at a given capacity.

4. **A Multiple Scanning Histogram Modification Example.** In this section we discuss the improvement made by using multiple scanning techniques in calculating the difference value histogram. Fig. 5 shows difference value histograms of the Lena image using the horizontal, vertical, and diagonal scanning techniques. For the Lena image, the vertical scanning technique has higher peaks for the difference values close to 0. Therefore, we first perform histogram modification using the vertical scanning technique. For demonstrative purposes, we will arbitrarily choose an EL of 3 for embedding.

Fig. 6 shows the Lena difference value histograms after performing the horizontal histogram shifting algorithm with an EL of 3. While all of the difference value histogram peaks around 0 have decreased, the peak values using the horizontal scanning technique are higher than those of the vertical scanning difference histogram of the original image for an embedding level greater than 3. As a result, the payload capacity is increased by first embedding using vertical scanning and then horizontal scanning rather than embedding using only one method at an increased embedding level.

We perform histogram shifting again this time using horizontal scanning with an EL of 3 followed by diagonal scanning with an EL of 3. At this point we have embedded 303,076 bits of data for a payload capacity of 1.16 bpp. This exceeds the theoretical maximum payload of Zhao et al.'s algorithm. However, it is unclear by solely examining the difference value histograms which order of scanning techniques should be chosen and what

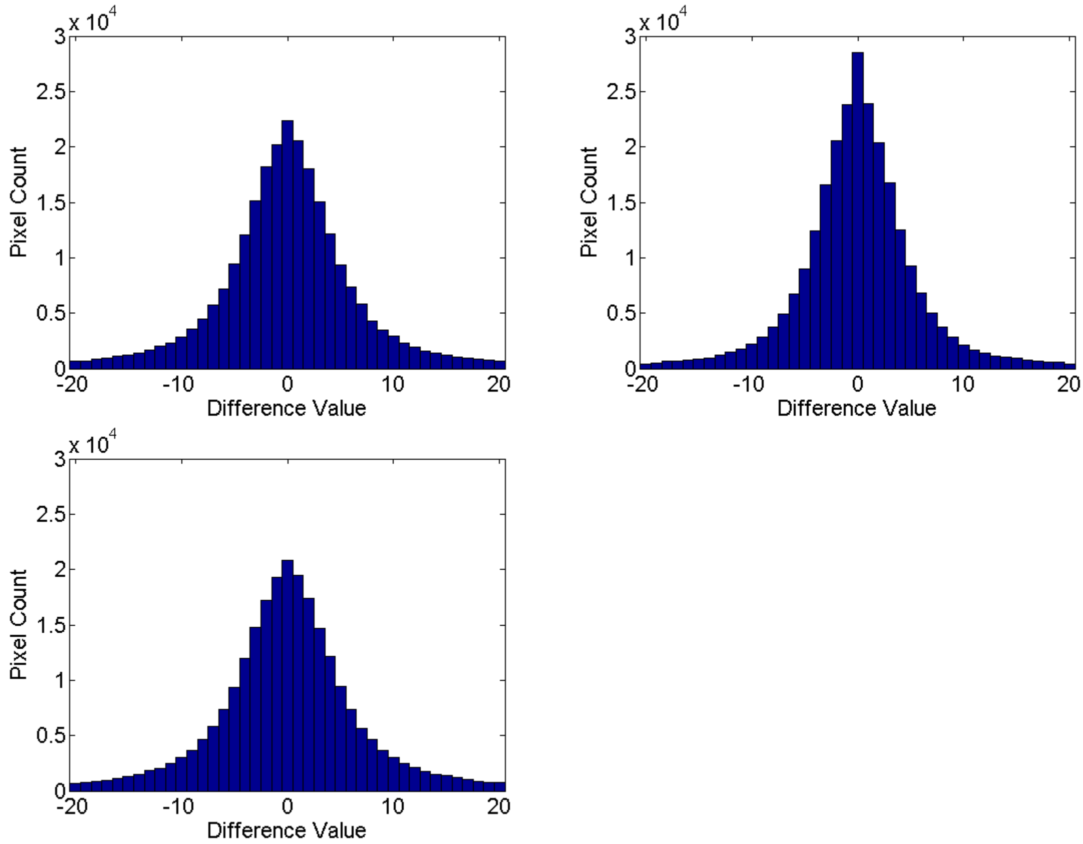

FIGURE 5. Difference value histogram of the Lena image using (a) horizontal scanning (b) vertical scanning, and (c) diagonal scanning

embedding level should be used. The next section presents an algorithm that provides a high capacity embedding order.

5. **An Iterative Algorithm for Determining Embedding EL and Scan Order.** A typical image can have over 40 histogram shifting operations applied to it before a reduction in payload capacity is experienced. This means that there are approximately 403 = 64,000 combinations of histogram shifts that can be applied to the image before the payload capacity is reduced. Since it is computationally expensive to perform such a large number of histogram shifts, a brute force method of determining the histogram shift and EL order is undesirable. Therefore, we propose an iterative algorithm that aims to maximize the payload capacity while maintaining low computational complexity.

The proposed algorithm for optimizing the histogram shift and EL order is an application of the hill climbing technique. We use this technique to maximize the payload capacity that is gained at each step while minimizing the increase in the cover image noise. We store a list of proposed methods. A proposed method is defined as a sequence of pairs of scanning methods and embedding levels. For each proposed method, we store the capacity and PSNR of the produced cover image at each EL.

We begin by initializing the proposed methods as the three scanning methods at EL = 0. The payload and PSNR of the produced cover images are saved with the corresponding method. The active method is defined as the method with the highest payload to PSNR ratio. The first active method is set as the maximum payload of the proposed methods. Then we perform the following iterative operation:

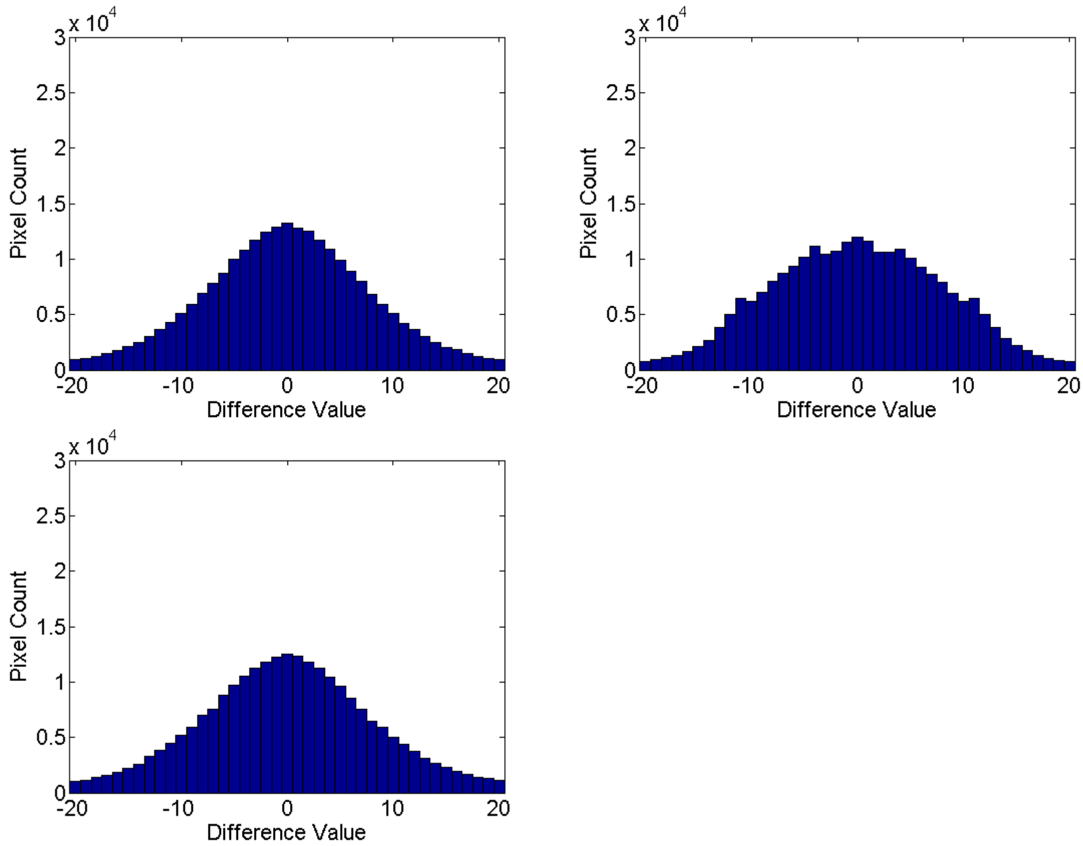

FIGURE 6. Difference value histograms of the once embedded Lena image using (a) horizontal scanning (b) vertical scanning and (c) diagonal scanning

1. Apply each histogram shifting using EL = 0 to the cover images of the current method using each scanning technique. We call the method used in this step the current embedding method.
2. Compare the payload capacity and PSNR to the proposed method.
3. If the payload capacity of a proposed method is less than that of the current embedding method, then increase the EL of the proposed method's last scanning technique until it is greater than that of the current embedding method or until the proposed method technique decreases.
4. If the current embedding method has a greater payload capacity than all of the proposed methods for the PSNR of the cover image of the current embedding method, then the next step is defined as the current embedding method. Otherwise, the next step is defined as the proposed method with the highest payload capacity
5. Repeat steps 2-4 for each of the scanning techniques.
6. If the next step is one of the proposed methods, then set the current method to the proposed method, changing the EL of the last step to the EL that was set in step 2b. Otherwise, append the next method to the current method.
7. Repeat steps 1-6 until the desired payload capacity is reached or there is no increase in the payload capacity.

6. **Experimental Results.** We perform our proposed algorithm on six test images shown in Fig. 7. Our proposed algorithm is compared with the single scanning method algorithm. Payloads are pseudorandomly generated such that the maximum number of

bits is embedded in each image. The PSNR and payload capacity are compared for each algorithm.

FIGURE 7. Test images: (a) Lena, (b) Airplane, (c) Car, (d) Peppers, (e), Barbara, and (f) Goldhill

Fig. 8 shows the payload capacity of the test images and the PSNR of the produced cover images using Zhao et al.'s algorithm and our proposed algorithm. Since the single horizontal scanning method is always in the solution scanning space of our EL and scan order algorithm, our algorithm always produces at least as much storage capacity for a given PSNR value. The proposed algorithm has a substantially improved maximum payload capacity. While Zhao et al's algorithm has a theoretical maximum of 1 bpp, our algorithm consistently exceeds this boundary, reaching higher than 2.5 bpp in the Airplane image.

Our algorithm has a higher maximum capacity at the 30dB PSNR level as well as a higher overall maximum payload capacity for every test image. At the 30 dB PSNR level, which is considered to be the limit of human perception, our algorithm has an average maximum payload capacity of 1.1179 bpp, compared to 0.8530 with Zhao et al.'s algorithm, a 31% improvement. The overall maximum average maximum payload capacity of our algorithm is 1.9233, compared to 0.8706, a 120% improvement.

7. **Conclusions.** We have proposed an improved histogram modification reversible data hiding algorithm using multiple scanning techniques. Our technique takes advantage of the fact that the central peaks of the difference value histograms produced by additional scanning techniques are higher than the peaks of increasing the EL of the single horizontal scanning technique. This allows for an image to have more data embedded within it. We have also introduced an algorithm for determining an appropriate scan order and EL. This produces payloads that are at least as high as Zhao et al.'s algorithm at any noise level and are higher in all of the test images at 30 dB.

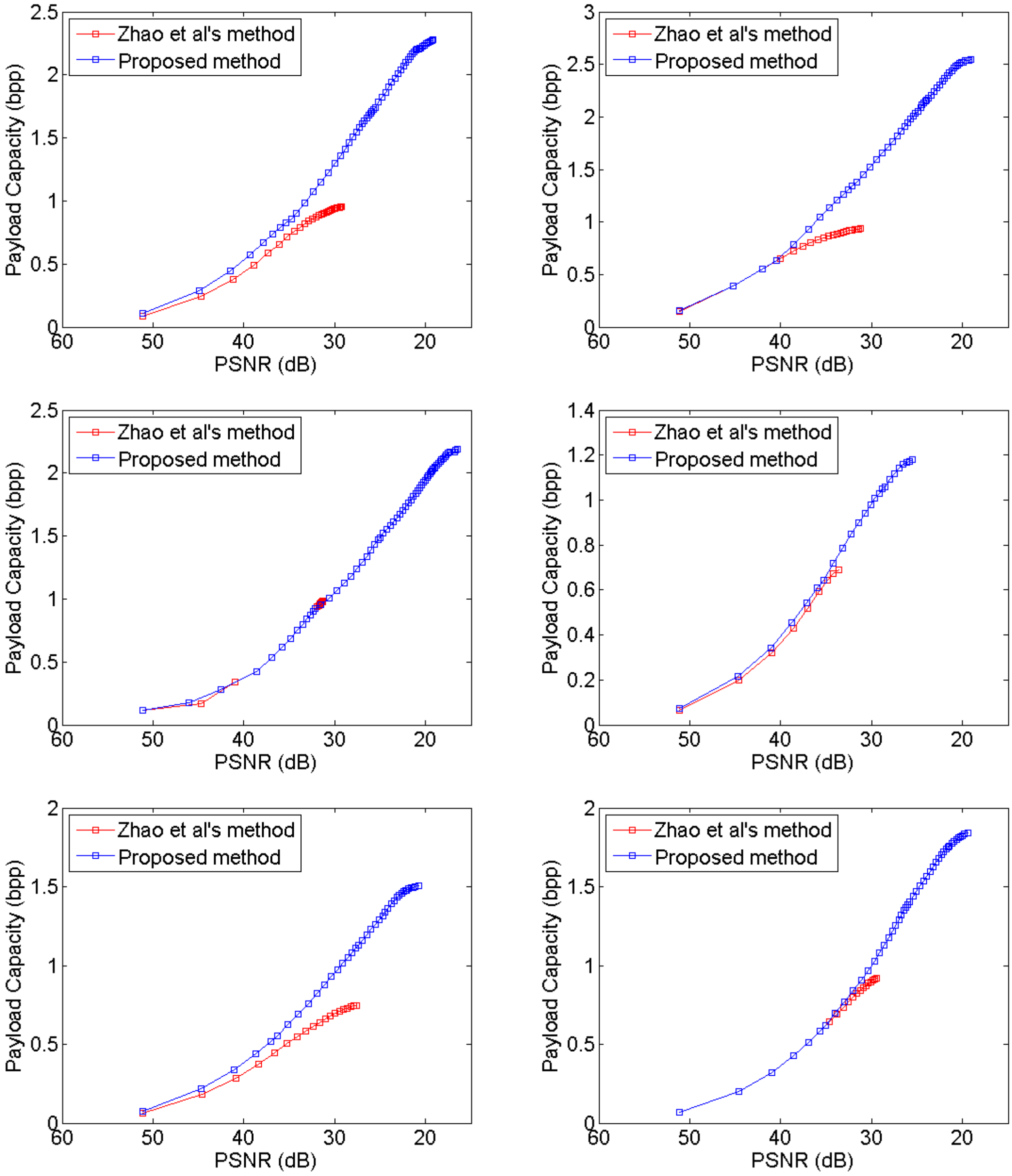

FIGURE 8. The noise of the cover images (PSNR) plotted with the payload capacity of Zhao's algorithm and the proposed algorithm of the following images (a) Lena, (b) Airplane, (c) Car, (d) Peppers, (e), Barbara, and (f) Goldhill

REFERENCES

[1] F. Y. Shih, *Multimedia Security: Watermarking, Steganography, and Forensics*, CRC Press, Boca Raton, FL, 2013.

[2] J. Tian, Reversible data hiding using a difference expansion, *IEEE Trans. Circuits & Systems for Video Technology*, vol. 13, no. 8, pp. 890-896, 2003.

[3] A. M. Alattar, Reversible watermark using the difference expansion of a generalized integer transform, *IEEE Trans. Image Processing*, vol. 13, no. 8, pp. 1147-1156, 2004.

[4] H. J. Kim, V. Sachnev, Y. Q. Shi, J. Nam, and H. G. Choo, A novel difference expansion transform for reversible data embedding, *IEEE Trans. Information Forensics and Security*, vol. 3, no. 3, pp.456-465, 2008.

Table 1. Comparison of Payload Capacity

Image	Zhao et als Method		Our Proposed Method	
	Max Capacity PSNR<30 (bpp)	Max Capacity Overall (bpp)	Max Capacity PSNR<30 (bpp)	Max Capacity Overall (bpp)
Lena	0.9367	0.9516	1.2952	2.2786
Airplane	0.9344	0.9344	1.5227	2.5444
Car	0.9806	0.9806	1.0098	2.1889
Peppers	0.6903	0.6903	0.9805	1.1800
Barbara	0.6794	0.7478	0.9296	1.5062
Goldhill	0.8964	0.9190	0.9694	1.8416
Average	0.8530	0.8706	1.1179	1.9233

[5] S. W. Weng, Y. Zhao, and J. S. Pan, A novel reversible data hiding scheme, *International Journal of Innovative Computing, Information and Control*, vol. 4, no. 2, pp. 351-358, 2008.

[6] Z. Ni, Y. Q. Shi, N. Ansari, and W. Su, Reversible data hiding, *IEEE Trans. Circuits and Systems for Video Technology*, vol. 16, no. 3, pp. 354-361, 2006.

[7] L. Kamstra, and H. Heijmans, Reversible data embedding into images using wavelet techniques and sorting, *IEEE Trans. Image Processing*, vol. 14, no. 12, pp. 2082-2090, 2005.

[8] D. Xiao, and F. Y. Shih, A reversible image authentication scheme based on chaotic fragile watermark, *International Journal of Innovative Computing, Information and Control*, vol. 6, no. 10, pp. 4731-4742, Oct. 2010.

[9] Z. H. Wang, M. T. Sun, and C. C. Chang, Reversible data-hiding techniques for digital images, *Multimedia Security: Watermarking, Steganography, and Forensics*, CRC Press, Boca Raton, FL, pp.115-139, 2013.

[10] J. D. Lee, Y. H. Chiou, and J. M. Guo, Reversible data hiding based on histogram modification of SMVQ indices, *IEEE Trans. Information Forensics and Security*, vol. 5, no. 4, pp. 638-648, 2010.

[11] Z. Zhao, H. Luo, Z. Lu, and J. Pan, Reversible data hiding based on multilevel histogram modification and sequential recovery, *AEU - International Journal of Electronics and Communications*, vol. 65, no. 10, pp. 814-826, 2011.

[12] C. Y. Yang, W. C. Hu, and C. H. Lin, Reversible data hiding by coefficient-bias algorithm, *Journal of Information Hiding and Multimedia Signal Processing*, vol. 1, no. 2, pp.91-100, 2010.