

High capacity lossless data hiding based on histogram modification

M. Fallahpour $^{\rm a)}$ and M. H. Sedaaghi $^{\rm b)}$

Faculty of Electrical Engineering Sahand University of Technology, Tabriz, Iran a) falahpour@unistar-micro.com, fallahpour@gmail.com b) Sadaaphi@gut ap in Sadaaphi@gija.osd guth ar

b) Sedaaghi@sut.ac.ir, Sedaaghi@aiia.csd.auth.gr

Abstract: This letter introduces a lossless reversible data hiding method. The original image is recovered without any distortion from the marked image following the extraction of the hidden data. The main algorithm is based on the relocation of zeros (or minima) and peaks of the histograms of the blocks of the original image to embed the data. It leads to the modification of the grey values of some pixels. It can embed more data than many of the existing reversible data hiding algorithms.

The PSNR of the proposed method is better than many of the existing reversible data hiding techniques. Some of the huge experimental results are presented to prove its validity.

Keywords: watermarking, histogram modification, reversible data hiding

Classification: Science and engineering for electronics

References

- Ni Zhicheng, Y. Q. Shi, N. Ansari, and W. Su, "Reversible data hiding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 3, pp. 354–362, March 2006.
- [2] M. U. Celik, G. Sharma, A. M. Tekalp, and E. Saber, "Reversible data hiding," *Proc. IEEE Int. Conf. Image Processing*, vol. 2, pp. 157–160, Sept. 2002.
- [3] A. Leest, M. Veen, and F. Bruekers, "Reversible image watermarking," *Proc. IEEE ICIP*, vol. 2, pp. 731–734, Sept. 2003.
- [4] J. Tian, "Reversible data embedding using a difference expansion," IEEE Trans. Circuits Syst. Video Technol., pp. 890–896, Aug. 2003.
- [5] M. Fallahpour and M. H. Sedaaghi, "Lossless adaptive data hiding," WSEAS Trans. Signal Processing, vol. 2, no. 6, pp. 909–913, June 2006.
- [6] G. Xuan, Y. Q. Shi, and Z. Ni, "Lossless data hiding using integer wavelet transform and spread spectrum," *IEEE Int. Workshop on Multimedia Signal Processing*, Siena, Italy, Sept. 2004.
- [7] Y. Q. Shi, Z. Ni, D. Zou, C. Liang, and G. Xuan, "Lossless data hiding: Fundamentals, algorithms and applications," *Proc. IEEE Int. Symp. Circuits Syst.*, vol. II, pp. 33–36, Vancouver, Canada, May 2004.
- [8] Y. Q. Shi, "Reversible data hiding," Proc. Int. Workshop on Digital Wa-





termarking, Seoul, Korea, Oct. 1 to Nov. 2, 2004.

1 Introduction

Invisibility of the embedded data in data hiding is a critical point for most applications such as medical images and law enforcement. A variety of datahiding algorithms have been proposed for different applications. Ni et. al. have proposed a histogram-manipulation-based lossless data hiding algorithm [1], we refer to their paper frequently. A generalization of the LSB modification suggested by Celik et. al. [2], presents a high capacity, low distortion reversible data hiding technique. Leest et al. [3] reported a reversible image watermarking algorithm using the gaps technique. Tian [4] embeds data using the difference expansion technique which results in one of the best reversible data hiding method among all the existing reversible data hiding techniques. These algorithms are applied in spatial domain.

Our group [5] have also presented an adaptive lossless data hiding method based on histogram manipulation which is more efficient than other existing methods in spatial domain. Lots of researches (related to data hiding) have been performed on transform domain among which IWT exists. Xuan et al. [6] reported the reversible data hiding algorithms carried out in the IWT domain. This algorithm applies spread-spectrum technique to embed data in high frequency IWT coefficients. For a survey, the readers may refer to [7, 8].

In this letter, we propose a new reversible data embedding algorithm where a huge amount of data (5 K-250 K bits) can be embedded in a grey-scale image of size 512×512 . We consider the zeros (or minima) of the blocks of the original image. Then, we slightly modify the grey value of some of the pixels of different blocks belonging to the original image. This technique can be applied to virtually all types of images.

The rest of the paper is organized as follows. The proposed algorithm is introduced in Section 2. Some experimental results and conclusions are presented in Sections 3 and 4, respectively.

2 The proposed method

The proposed method is explained in the two following subsections.

2.1 Embedding

The embedding algorithm is as follows.

1) The image is first divided into series of blocks (denoted by N_b , e.g. 4, 16, etc.). Subsequently, the following steps (2-4) are iterated for each block.

2) For a given n, the same amount of the pairs of peak and zero of the histogram is found (for current block Pi is i^{th} peak and Zi is i^{th} zero).

3) The following iteration is looped n times for i =1: n.





4) For pair (Pi, Zi):

a) If Pi>Zi, then the whole block is scanned and the gray values of the pixels between Zi +1 and Pi are reduced by one (shifting the range of the histogram [Zi +1, Pi] by 1 to left). This causes a gap in gray value Pi. Afterward, the block is scanned and the gray value of the pixels with gray value of Pi-1 are increased by one if the corresponding bit of the data (to be embedded) is one, otherwise they will not be modified.

b) If Zi>Pi, the whole block is scanned and the gray values of the pixels between Pi +1 and Zi-1 are increased by one. This causes a gap in gray value Pi +1. Then block is scanned and the gray values of the pixels with gray value of Pi are increased by one if the corresponding bit of the data (to be embedded) is one, otherwise they will not be altered.

The gray value of the zero points and the peak points will be treated as side information that needs to be transmitted to the receiving side for data retrieval.

2.2 Detection

The detection algorithm is as follows.

- 1) Firstly, the image is divided into series of blocks (denoted by N_b). Then steps 2-3 are repeated for each block.
- 2) The following iteration is done n times for i = 1: n.
- 3) For pair (Pi, Zi):

a) If Pi>Zi, the whole block is scanned. The pixels with gray value Pi indicates that the embedded data bit is 1 and they should not be modified. Otherwise, if it is equal to Pi-1, it clarifies that the embedded data bit is 0. In this case, their gray values have to be increased by 1. Later, the gray values of all pixels with gray values between Zi and Pi-2 need to be increased by one.

b) If Zi>Pi, the whole block is scanned. The pixels with gray value Pi indicates that the embedded data bit was 0 and they do not need to be modified. However, if it is equal to Pi +1, it clarifies that the embedded data bit is 1. Then, their gray values are reduced by 1. Therefore, the gray values of all pixels with gray value between Pi +2 and Zi are reduced by 1.

The shift of the peaks and zeros should not lead to losing the prior information about the location(s) of peaks and zeros. It is noteworthy that, in any case, we use the minima instead of zeros in the histogram if there is no sufficient number of zeros [1].

2.3 Improvement of capacity

In Ni algorithm a pair of peak and zero is used, P is peak of whole image and amount of peak points is h(P). These peak points are in whole image. Supposedly, image is divided into N_b blocks. Some of these peak points are





in block one, some of these are in block two and etc. Assume the amount of peak points in i^{th} block (i=1 to N_b) is $h_i(p)$.

In algorithm discussed on this paper, the image is divided into N_b blocks. The peak corresponding to each block can be different from the others because of using embedding algorithm separately for each block. Assume p_i is peak point of i^{th} (i=1 to N_b) block and the amount of peak points in i^{th} (i=1 to N_b) block is $h(p_i)$.

$$h_i(p) \le h(p_i)$$

Because of the assumption p_i is peak point of i^{th} block.

$$\sum_{i=1}^{N_b} h_i(p) \le \sum_{i=1}^{N_b} h(p_i)$$
$$h(p) \le \sum_{i=1}^{N_b} h(p_i)$$

Therefore, this shows that using embedding algorithm seperately for each block of image improve the size of embedding data.

PSNR of blocks may be higher or lower than PSNR of Ni method for whole image. Probabaly PSNR of whole image after using algorithme of this paper is lower than PSNR obtained by Ni method. In this case P, can be used instead of Pi, in blocks with lower PSNR than PSNR of Ni method. Hence, PSNR of the whole image is higher than that in Ni method.

3 Experimental results and evaluations

The results tested on different images, have confirmed the superiority of the proposed method. This method increases the capacity to a level more than what obtained by Ni et. al. [1] where the resulting PSNR is still high enough. Massive experimental results prove that the embedded data remains invisible and there is no visual distortion.

Fig. 1 shows comparison between Ni et. al. [1] and purposed method. In Lena's image, the two peaks have capability of embedding 5460 bits and PSNR= 48.2 dB while two peaks of each four blocks of the same image will have the embedding capacity of 2507+2815+1941+2308=9571 bits and PSNR= 47.3 dB.

Table I summarizes the experimental results obtained by our method. The results could even be better by increasing the number of blocks and pairs of peak and zero. Fig. 2 compares the performance in terms of PSNR and payload (bpp: bits per pixel) between the proposed algorithm and the method proposed by Ni et. al. [1] for the Lena, Airplane, Pepper, Sailboat images. We realize that the capacity is increased while PSNR is decreased by augmenting the number of pairs of peak and zero. Fig. 2 confirms the better performance of our method as we have employed 4 (16) blocks instead of only one block (Ni's method).

The only minor inconvenience is that the increasing number of the blocks leads to more complexity. It also has an upper limit, as the effect of the







Fig. 1. The Marked image and its histogram (a) Whole image (PSNR=48.2) (b) 1st block (PSNR=48.33);
(c) 2nd block (PSNR=46.36) (d) 3rd block (PSNR=50.18) (e) 4th block (PSNR=45.7).

Table I.	PSNR and capacity of the proposed method com-
	pared with Ni's algorithm [1].

Image	Pure payload					PSNR of marked image(dB)		
	Ni's[1,3] method	The proposed method used 4 Blocks		The proposed method used 16 Blocks		Ni's [1,3]	This paper method	This paper method
	(bits)	bits	Improvement percent	bits	Improvement percent	method	used 4 blocks	used 16 blocks
Lena	5460	9571	75%	13868	154%	48.2	47.30	47.29
Airplane	16171	24421	51%	29250	81%	48.3	48.54	48.56
Tiffany	8782	15245	74%	22011	151%	48.2	48.3	48.4
Jet	59979	99099	65%	137141	129%	48.7	48.8	48.7
Baboon	5421	5892	9%	7874	46%	48.2	48.35	47.12
Pepper	5449	9499	74%	13185	142%	48.2	48.06	47.7
Sailboat	7301	9039	24%	15231	108%	48.2	51.50	47.84
Boat	7301	12018	65%	15579	113%	48.2	47.85	46.73







Fig. 2. Capacity versus PSNR (a) Lena; (b) Airplane (c) Pepper (d) Sailboat

blocking will appear on the quality of the host image. It is observed that the proposed technique has achieved the high enough bound of the PSNR with a quite large data embedding capacity.

4 Conclusion

We employ the pairs of zero and peak of the blocks of the image instead of the corresponding pairs of the whole image. One of the advantages of our algorithm is its simplicity in implementation. That will be useful for a reversible algorithm in spatial domain. In fact, this algorithm has been successfully applied to many frequently used images such as medical, texture and military images. The method we offer increases the capacity where the resulting PSNR is still high enough. Our reversible data hiding technique is able to embed about 5 k- 250 k bits into a 512×512 grayscale image and has very good PSNR. The gray values in military and medical images are mostly concentrated around some gray values and the rest are not occupied. That is why our algorithm has superior results for data hiding in military and medical images. Our method has advantage to the method reported in [1] where their algorithms used to be considered as a record in spatial domain.

