Lossless Data Hiding Using Integer Wavelet Transform and Threshold Embedding Technique

Guorong Xuan[1], Yun Q. Shi[2], Chengyun Yang[1], Yizhan Zheng[1], Dekun Zou[2], Peiqi Chai[1]
[1] *Tongji University, Shanghai, China (grxuan@public1.sta.net.cn)*
[2] *New Jersey Institute of Technology, Newark, New Jersey, USA (shi@njit.edu)*

Abstract

This pape[1]r presents a new lossless data hiding method for digital images using integer wavelet transform and threshold embedding technique. Data are embedded into the least significant bit-plane (LSB) of high frequency CDF (2,2) integer wavelet coefficients whose magnitudes are smaller than a certain predefined threshold. Histogram modification is applied as a preprocessing to prevent overflow/underflow. Experimental results show that this scheme outperforms the prior arts in terms of a larger payload (at the same PSNR) or a higher PSNR (at the same payload).

1. Introduction

The development of information technologies makes it convenient for people to transmit mass data through Internet. However, it also provides vast opportunities for hackers to steal valuable information. Therefore, security becomes an important issue. Digital data hiding can hide sensitive information into multimedia for covert communications. Most multimedia data hiding techniques will distort the cover media in order to insert the additional information. Although the distortion is often small and imperceptible to human visual systems (HVS), the irreversibility is not admissible to some sensitive applications, such as legal and medical imaging. For these applications, lossless data hiding is desired to extract the embedded data as well as recover the original host signal. About 20 lossless data hiding methods have been developed. For a survey, readers are referred to [1,2]. Ni et al. have proposed a histogram-manipulation based lossless data hiding scheme [3]. Leest et al. [4] proposed a reversible image watermarking algorithm using the gaps technique. Tian [5] embeds data using the difference expansion technique and results in one of the best reversible data hiding method among all the existing reversible data hiding techniques. Xuan et al. proposed the reversible data hiding algorithms carried out in the integer wavelet transform (IWT) domain. One method [6] losslessly compresses one or more than one middle bit-planes to save space for data embedding. The bookkeeping data are also embedded as overhead. Another [7] applies spread-spectrum technique to embed data in high frequency IWT coefficients. In this paper, a novel lossless data hiding method for digital images using integer wavelet transform and threshold embedding technique is proposed. The proposed lossless data hiding technique is rather simple and, yet, outperforms the prior arts. Both theoretical analysis and experimental results demonstrate the superiority of the proposed technique.

The rest of the paper is organized as follows. A brief introduction to wavelet transform and histogram modification is provided in Section 2. Section 3 describes the idea of threshold embedding. The proposed data hiding system is presented in Section 4. Some experimental results and performance analysis are presented in Section 5. The conclusion is drawn in Section 6.

2. Wavelet Decomposition and Histogram Modification

We propose to use the CDF (Cohen-Daubechies-Fauraue) (2,2) integer wavelet transform, adopted by JPEG2000 for image lossless compression, to obtain the wavelet coefficients. Because of what is called frequency mask, the data embedded into in the high frequency subbands, HL, LH and HH, will have less visible artifact to human eyes.

After data are embedded into some high frequency IWT coefficients, it is possible that after inverse integer wavelet transform, the grayscale values of some pixels in the marked image may exceed the upper bound (255 for an eight-bit grayscale image) and/or the lower bound (0 for an eight-bit grayscale image). This phenomenon is called overflow/underflow. In order to

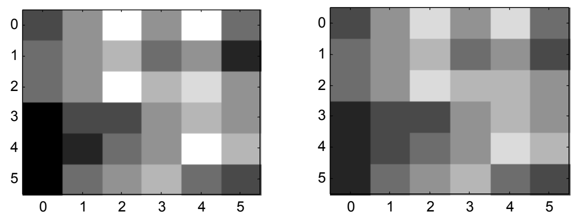
[1] This research is supported partly by National Natural Science Foundation of China (NSFC) on the project "The Research of Theory and Key Technology of Lossless Data Hiding (90304017)", and by New Jersey Commission of Science and Technology via New Jersey Center of Wireless Networking and Internet Security (NJWINS).

0-7803-9332-5/05/$20.00 ©2005 IEEE

prevent the overflow and underflow, histogram modification is applied to narrow down the histogram from both sides.

In order to illustrate the histogram narrow down process, we use the following simplified example, where the size of an original image is 6x6 with $8=2^3$ grayscales (6x6x3) as shown in Fig. 1.

From Fig. 1 and Table 1, we can see that the range of the modified histogram now is from 1-6 instead of 0-7, i.e., no pixel assumes gray scales 0 and 7. After modification, grayscale 1 is merged into gray scale 2. Grayscale 0 becomes grayscale 1. In the same way, grayscale 6 is merged into grayscale 5. Grayscale 7 becomes grayscale 6. Histogram before and after modification, and bookkeeping information are shown in Table 1 and Table 2, respectively.



(a) Original image (b) After histogram modification

2	4	7	4	7	3
3	4	5	3	4	1
3	4	7	5	6	4
0	2	2	4	5	4
0	1	3	4	7	5
0	3	4	5	3	2

(c) Original image Data

2	4	6	4	6	3
3	4	5	3	4	2
3	4	6	5	5	4
1	2	2	4	5	4
1	2	3	4	6	5
1	3	4	5	3	2

(d) Modified image data

Fig.1 Illustrative example of histogram modification.

Table 1. Histogram data before and after modification.

Grayscale value	0	1	2	3	4	5	6	7
Pixel no. before mod.	3	2	4	7	10	5	1	4
Pixel no. in mod.	3	0	6	7	10	6	0	4
Pixel no. after mod.	0	3	6	7	10	6	4	0

Table 2. Bookkeeping information.

For image (6×6×3), the histogram is narrowed down 1 gray scale for both sides. G=2, G/2=1. The total bits length is 37 bits.

S=the total book-keeping bit length 37 bits (00100101) + compressed number of gray scale 2 (010) +
the first histogram from left hand side grayscale "1" (001) +record length 6 (0110) + scan sequence (101101)+
the first histogram from right hand side grayscale "6" (110) + record length 6 (0110) + scan sequence (110111)
S=[00100101 010 001 0110 101101 110 0110 110111]

The left hand side record bits with its left neighbor grayscale (101101) in Table 2 shows that both second and fifth "2" by scanning (<x=5,y=1>,<x=1,y=4>) in Fig. 1 (d) are "1" in Fig. 1(c) originally. Also the right hand side record bits with its right neighbor grayscale (110111) in Table 2 shows that the third "5" by scanning (<x=4,y=2>) in Fig. 1(d) is "6" in Fig. 1(c) originally.

Though this example is simple, the histogram modification algorithm illustrated here can be applied to image of large size efficiently in terms of less computation and smaller amount of bookkeeping data.

This efficient histogram modification algorithm has been automatically and successfully applied to all of the 1096 images in the CorelDraw image database [8] when we tested our lossless data hiding algorithms in [6,7]. We have used it in this work extensively as well.

3. Threshold Embedding

Threshold embedding, another key technology of this scheme, is brand new in this paper. A threshold value *T* is predefined. Different embedding rules apply to high frequency wavelet coefficients according to the absolute value of the coefficient is smaller than *T*, equal to *T*, or larger than *T*. By high frequency wavelet coefficients, it means the IWT coefficients of high frequency subbands, i.e., the coefficients in the HL, LH, and HH subbands.

3.1. Data hiding

To embed data into a high frequency coefficient x, the absolute value of the coefficient is compared with T. If $|x| < T$, the coefficient value is doubled and the new LSB is replaced with an information bit. Equivalently, the binary representation of the coefficient value is shifted towards left by one bit, and the to-be-embedded bit is appended as the right-most bit. The resultant coefficient is denoted by x'. Otherwise, if $x \geq T$, the coefficient will be added by T, if $x \leq -T$, the coefficient will be subtracted by (T-1), and no bit is embedded into this coefficient. These rules can be summarized in Formula (1).

$$x' = \begin{cases} 2 \bullet x + b, & if\ |x| < T \\ x + T, & if\ x \geq T \\ x - (T-1), & if\ x \leq -T \end{cases} \qquad (1)$$

3.2. Hidden data extraction and original image restoration

In the data extraction stage, the IWT coefficients of the marked image are obtained by applying IWT. For a coefficient, if it is less then 2T and larger than (-2T+1), the LSB of this coefficient is the bit embedded into this coefficient. Otherwise, we jump to the next coefficient since the current coefficient has no hidden bit in it. Besides hidden data extraction, the original cover image should be able to be recovered. Concretely, each high frequency coefficient can be restored to its original value by applying Formula (2).

$$x = \begin{cases} \lfloor x'/2 \rfloor, & if \ -2T+1 < x' < 2T \\ x' - T, & if \ \ x' \geq 2T \\ x' + T - 1, & if \ \ x' \leq -2T+1 \end{cases} \qquad (2)$$

where $\lfloor y \rfloor$ takes the largest integer value that is smaller than y.

Table 3 provides an example to hide two bits: 0, followed by 1 into four coefficients [3, 5, 9, -7], where T is set to be 6. It is easy to verify the embedding rule, data extraction rule and original coefficients recovery.

Table 3 Embed bits 0, and 1 into coefficients [3,5,9,-7] with T=6.

Order	Original value	Greater than T?	Hidden data [0,1]	Marked value
1	3	No	0	6
2	5	No	1	11
3	9	Yes	-	15
4	-7	Yes	-	-12

4. Proposed Data Hiding System

The proposed lossless data hiding scheme embeds data into the first level high frequency subbands of images, namely, HL_1, LH_1 and HH_1.

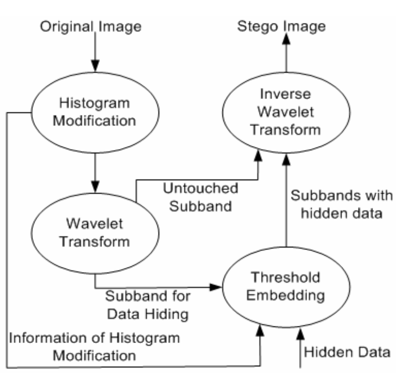


Fig. 2 Flowchart of the proposed lossless data hiding using threshold embedding.

Histogram modification is performed prior to data embedding to ensure no overflow/underflow will take place. The bookkeeping data of histogram modification and the payload are to be embedded into the high frequency IWT coefficients. The stego-image carrying hidden data will be obtained after inverse integer wavelet transform. Fig. 2 is the flowchart of the proposed threshold embedding data hiding.

Fig. 3 is the flowchart for hidden data extraction and original cover image recovery.

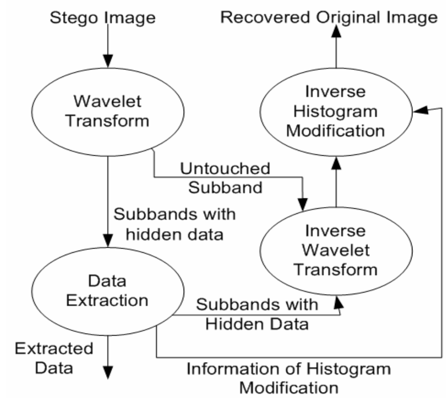


Fig. 3 Flowchart of proposed hidden data extraction and cover image recovery using threshold embedding.

5. Discussions and Examples

For most of images, the distribution of high frequency coefficients of integer wavelet transform obeys a Laplacian-like distribution. Here, the high frequency coefficients mean the coefficients in the HL, LH or, HH subband. Most high frequency integer wavelet transform coefficients are very small in magnitude. Since, the high frequency coefficients in three high frequency subbands consist 75% of all IWT coefficients, and one bit can possibly be embedded into a coefficient. Therefore, the highest payload is 0.75 bit per pixel (bpp). For Lena image, if threshold T is set to be 8, the payload is 0.68 bpp. It shows that over 90% coefficients in the high frequency subbands are used for data hiding (0.68/0.75 = 0.9067). Table2 lists the payload of Lena image under different Thresholds.

Table 4 Threshold v.s. payload for Lena image.

T = 4	T = 6	T = 8
0.50 bpp	0.62 bpp	0.68 bpp

Clearly, the larger the threshold T, the higher the payload will be. However, on the other hand, the larger the threshold T, the lower the PSNR (peak signal noise ratio) will be. The large threshold means the strong embedding strength. These are general true for all data hiding methods. That is, the payload and PSNR are contending one another and the embedding strength heavily influences these two parameters. Fig. 4 compares the performance in terms of PSNR and payload between the proposed algorithm and some other existing lossless data hiding methods. There, the blue curve marked by “Threshold Embedding Based on IWT” is the method proposed in this paper. The orange curve marked by “Bitplane Compression” is the method reported in [6]. The pink curve marked by “Difference Expansion” is the method reported in [5]. The black curve marked by “Spread Spectrum” is reported in [7]. It can be observed that the proposed method has the highest PSNR under the same payload, or in other words have the largest payload under the same PSNR of the stego-image versus the original image. Specifically, when the payload is set at 0.5 bpp, the proposed lossless data hiding algorithm can achieve about 4 dB increase in the PSNR.

Note that here only a single-time data embedding is considered.

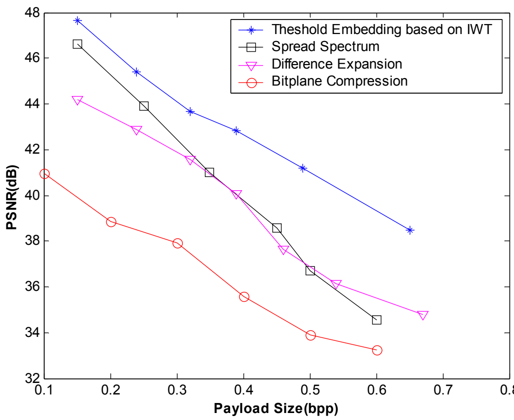


Fig. 4 Payload v.s. PSNR for several recently developed different lossless data hiding methods.

To further improve the payload, multi-time embedding technique can be adopted. Specifically, the stego-images can be treated as an original image. By applying the same embedding procedure to the stego-images, the payload can be increased. The drawback is the decrease of PSNR, in other words, deterioration of image quality. Fig.5 displays the Lena image with multi-time embedding. In Fig.5 (d), the visual quality of the marked image is not good and visual artifacts can be observed.

6. Conclusion

In this paper, a novel lossless data hiding method for digital images using integer wavelet transform and threshold embedding technique is proposed. The original cover image can be recovered losslessly if the stego-image has not been lossily processed. The payload and the visual quality of the stego-image is the best among the existing lossless data hiding methods.

(a)Original

(b) 0.7 bpp, PSNR: 37.46
(One time embedding)

(c) 1.3 bpp, PSNR: 31.1
(Two times embedding)

(d) 1.7 bpp, PSNR: 25.7
(Three times embedding)

Fig.5 Multi-time embedding.

7. References

[1] Y. Q. Shi, Z. Ni, D. Zou, C. Liang and G. Xuan, “Lossless data hiding: Fundamentals, algorithms and applications,” *Proceedings of IEEE International Symposium on Circuits and Systems*, vol. II, pp. 33-36, Vancouver, Canada, May 2004.

[2] Y. Q. Shi, “Reversible data hiding,” *Proceedings of International Workshop on Digital Watermarking*, Seoul, Korea, Oct. 1 to Nov. 2, 2004.

[3] Z. Ni, Y. Q. Shi, N. Ansari and W. Su, “Reversible data hiding,” *Proceedings of IEEE International Symposium on Circuits and Systems,* vol. 2, pp. 912-915, Bangkok, Thailand, May 2003.

[4] A. Leest, M. Veen, and F. Bruekers, “Reversible image watermarking,” *Proc. of IEEE ICIP*, vol.2, pp.731-734, Sept. 2003.

[5] J. Tian, “Reversible data embedding using a difference expansion,” *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 890-896, Aug. 2003.

[6] G. Xuan, J. Zhu, J. Chen, Y. Q. Shi, Z. Ni and W. Su, “Distortionless data hiding based on integer wavelet transform,” *IEE Electronics Letters*, Dec. 2002, pp. 1646-1648.

[7] G. Xuan, Y. Q. Shi, Z. Ni, “Lossless data hiding using integer wavelet transform and spread spectrum,” *IEEE International Workshop on Multimedia Signal Processing*, Siena, Italy, September 2004.

[8] www.corel.com