Using RZL Coding to Enhance Histogram-pair Based Image Reversible Data Hiding[1]

Xuefeng Tong[1], Guorong Xuan[1], Guangce Shen[1], Xiaoli Huan[1], Yun Q. Shi[2]

[1]Dept of Computer Science, Tongji University, Shanghai, China
[2]ECE, New Jersey Institute of Technology, Newark, New Jersey, USA
Guangce Shen<sgcskytt@163.com>

Abstract. An improvement of histogram-pair based image reversible data hiding by using RZL (Reverse Zero-run Length) coding is proposed in this paper. The pre-processing to compress data to a shortest one is usually adopted for raising the PSNR (Peak Signal to Noise Ratio) in data hiding. Recently, the disagreements appear that we can get better PSNR by using RZL coding after compression. We proved that our histogram-pair based image reversible data hiding is suitable to use RZL to improve the performance. The PSNR can be raised by using different RZL methods, different parameters, different embedded capacity and different images. It is hard to apply RZL to given original data with different lengths. We proposed a method to solve that by adding some 0s to the original data to form a complete block, and the RZL needs an attached mark for lossless recovery. In our experiments it has been shown that the PSNR of image with histogram-pair based reversible data hiding by using RZL is higher than that without using RZL as the embedding data rate is not high. Zhang et al.'s RZL is better than Wong et al.'s in most cases. The average PSNR gain is about 1 dB for five test images at different payloads with the RZL used in this paper.

Keywords: RZL (Reverse Zero-run Length) coding, Histogram-pair based image reversible data hiding, Decrement rate of the number of "1", Increment rate of length.

1 Introduction

Generally speaking, there are two common procedures that can be used to raise PSNR in image data hiding. The first procedure is to sharpen the distribution of the cover image's histogram, and the second is to shorten the data to be embedded. In

[1] This research is largely supported by Shanghai City Board of education scientific research innovation projects (12ZZ033) and National Natural Science Foundation of China (NSFC) on project (90304017).

addition, some particular caution needs to be taken care of. We will discuss them in detail below.

1.1 To sharpen the distribution of the cover image histogram

The first procedure to sharpen the distribution, in essential, is to decrease the entropy. For instance, we apply the transformation (e.g., prediction error, wavelet and DCT) before embedding, or we select the local area of the cover image with its histogram having sharp distribution for data embedding.

It is proved by Kalker and Willems [1] by using information theory, that the optimal embedding capacity for a memoryless binary source is $\rho_{rev}(p_0,\Delta) = H(\max(p_0-\Delta,1/2)) - H(p_0)$, where H( ) is entropy, p_0 ($p_0 \geq 1/2$) is probability of "0" in binary cover sequence $x=(x_1,x_2, \cdots;x_N)$, ρ_{rev} is optimal embedding rate $\rho_{rev}=L/N$, L is the length of the message $m=(m_1,m_2, \cdots;m_L)$, Δ ($0\leq \Delta \leq 1/2$, $\Delta=\delta/N$) is distortion, and δ is the modifications on average marked cover $y=(y_1,y_2, \cdots;y_N)$. Refer to Fig.1,

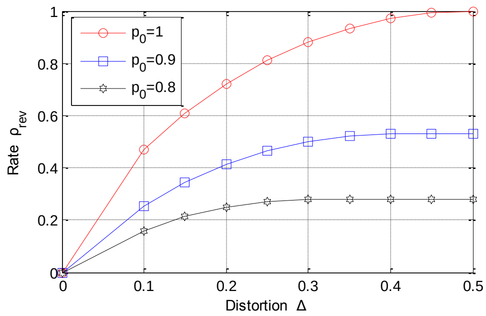

Fig. 1. Rate-distortion curves for p_0 = 1, 0.9 and 0.8 [1]

All-zero cover is just an extreme case of p_0=1 in binary cover sequence, which will produce the best environment for reversible data hiding. The histogram-pair scheme of image reversible data hiding [4] is just suitable to use RZL (Reverse Zero-run Length) because the all zero distribution cover is given in the histogram-pair based scheme.

1.2 To shorten the data to be embedded

The second procedure is to shorten the data to be embedded. But it is probably not the final optimal selection for the data pre-processing. Some pre-processing is recommended in recent years. After compression to shortest data, an RZL coding is used for decreasing the number of "1" in the data for further PSNR enhancement applied to the data hiding. An efficient RZL coding method for this case has been proposed by Wong et al. [2], and improved by Zhang et al. [3]. The improvement of histogram-pair based image reversible data hiding by using RZL coding is proposed in this paper.

Note that, to be short in discussion, in the rest of paper, we use Zhang to represent Zhang et al., similarly, Wong is also used to represent Wong et al.

1.3 Particular caution to be taken

In data hiding, an RZL procedure is adopted. That is, the special RZL types and parameters should be selected for different payloads and different images. Also, the RZL coding should be adapted for arbitrary length data cut off to realize lossless recovering.

1.4 Rest of the paper organized

The rest of the paper is organized as follows. Principles and formulas of two kinds of RZL coding are presented in Section 2. The improvement for arbitrary length RZL is given in Section 3. The histogram-pair based reversible data hiding with RZL coding is reported in Section 4. The conclusions are drawn in Section 5.

1.5 Notations

Some notations used in this paper are listed below.

m ---- Original data.
c ---- RZL code
λ ---- Increment rate of length $\lambda=L_1/L_0$（L_0 is length of *m*, L_1 is length of *c*）
σ ---- Decrement rate of the number of “1” $\sigma=S_1/S_0$
（S_0 is the number of binary “1” in *m*, S_1 is the number of binary“1” in *c*）
k ---- Length of block in original binary data $b=\{m_0, m_1, m_2, \ldots, m_{k-1}\}$for RZL
d ---- Digit from original binary data *b*, $d=m_0 2^0+m_1 2^1+m_2 2^2+\ldots+m_{k-1} 2^{k-1}$.

2 Principle and Formulas of Two Type RZL Coding

The RZL coding has been proposed by Wong [2], and an improving RZL method has been proposed by Zhang [3]. We do further improving by applying the RZL to histogram-pair based image lossless data hiding scheme. The principle and formulas comparison of RZL coding are listed in Tables 1~3.

Uniformly distributed binary random original data *m* can be transferred to RZL code *c*, in which the number of “1” will be less than that in original data, but *c* has length increasing. That is: $S_1< S_0$ with $L_1>L_0$, where S_0 and S_1 are the number of “1” in original data *m* and in the RZL code *c*, respectively. And L_0 and L_1 are the length of original data *m* and that of the RZL code *c* respectively. Both parameters σ and λ are adopted to describe the function of RZL, where σ=S_1/S_0 is the decrement rate of the number of "1" and λ=L_1/L_0 is the increment rate of length.

Above all,we know there are two factors of RZL coding:

1. Decrease the rate of 1, this factor is useful to raise the PSNR for Histogram-pair Based Image Reversible Data Hiding,
2. Increase the length of data, this factor is disadvantage to reduce the PSNR for Histogram-pair Based Image Reversible Data Hiding.

Whether raising or reducing PSNR by RZL depends on payload and image.

Table 1. Principle and formulas of Wong's RZL coding

Definition	Wong's RZL is to partition the original data *m* into blocks with equal length *k:* $\{m_0, m_1, m_2, \ldots, m_{k-1}\}$. The RZL coding is using a sequence of binary "0" with length *d* before a binary"1". The digit *d* is formed by length *k*, i.e., $d=m_0 2^0+m_1 2^1+m_2 2^2+...+m_{k-1}2^{k-1}$.
Example for Wong's RZL	● Example 1: Original data with k=3: m=011, where $d=\{d_1=3\}$. By using Wong's RZL, we have c=0001, and $(S_1=1) < (S_0=2)$ with $(L_1=4) > (L_0=3)$. ● Example 2: Original data with k=3: m=011101001={011}{101}{001}, where $d=\{d_1=3\}\{d_2=5\}\{d_3=1\}$. By using Wong's RZL, we have c={0001}{000001}{01}=000100000101, and $(S_1=3)<(S_0=5)$ with $(L_1=12) > (L_0=9)$.
Theoretical formula derivation	● We can prove the decrement rate of the number of "1" is $\sigma=S_1/S_0=2/k$. For uniformly distributed random data, "0" and "1" are equally probable. The number of "1" in m is $S_0=L_0/2$, the number of "1" in c (RZL code) is $S_1=L_0/k$, so we have $\sigma=S_1/S_0=2/k$. ● We can prove the increment rate of length is $\lambda=L_1/L_0=(2^k+1)/(2k)$. Because of uniformly distributed random data, "0" and "1" are equally probable. The RZL coding is using d number "0" sequence before a "1". So we have: $L_1=(L_0/k)([2^0+2^1+2^2+...+2^{(k-1)}]/2+1)$. The final increment rate of length is $\lambda=L_1/L_0=(2^k+1)/(2k)$ with simplification procedures.

Table 2. Principle and formulas of Zhang's RZL coding

Definition	Zhang's RZL is to partition the m to unequal length blocks in prefix "0" or "1". If we meet prefix "0", the block length is equal to one and RZL is using a sequence "0" with length 2^k. If we meet prefix "1", we use the following k bit for RZL coding. The RZL coding is using a sequence "0" with length of digit number d before a "1". The digit d is formed by a length k: $d=m_0 2^0+m_1 2^1+m_2 2^2+...+m_{k-1} 2^{k-1}$. Hence with a following "1", the block length is equal to k+1.
Example for Zhang's RZL	● Example 1: Original data with k=2: m=111, where $d=\{d_1=3\}$. By using Zhang's RZL, we have c=0001, and $(S_1=1)<(S_0=3)$with $(L_1=4)>(L_0=3)$. ● Example 2: Original data with k=2: m=0101110={0}{101}{110}, where $d=\{zero\}\{d_2=1\}\{d_3=2\}$. By using Zhang's RZL, we have c={0000}{01}{001}=000001001. And $(S_1=2) < (S_0=4)$ with $(L_1=9) > (L_0=7)$.
Theoretical formula derivation	● Because theoretically for random case, "0" and "1" are equally probable. The length of block prefix "0" in m is 1, and the length of block prefix "1" in m is (k+1). The length of a pair blocks with prefix"0" and "1" is (k+2). ● We can prove the decrement rate of the number of "1" number is $\sigma=S_1/S_0=2/(k+2)$. Because the number of "1" in m is $S_0=L_0/2$, the number of "1" in c (RZL code) is $S_1=L_0/(k+2)$, so we have $\sigma=S_1/S_0=2/(k+2)$. ● We can prove the increment rate of length is $\lambda=L_1/L_0=(2^{k+1}+2^k+1)/(2k+4)$. The block length of prefix "0" over a pair block length is $2^k/(k+2)$. The block length of prefix "1" over a pair block length is $([2^0+2^1+2^2+...+2^{(k-1)}]/2+1)/(k+2)$. Because of random assumption, "0" and "1" is equally probable. The increment rate of length is the sum of length of prefix "0" over a pair block and length of prefix "1" over a pair block: $\lambda=2^k/(k+2)+[(2^k-1)/2+1]/(k+2)$. The final increment rate of length is $\lambda=L_1/L_0= (2^{k+1}+2^k+1)/(2k+4)$ with simplification procedures.

We derive the formulas for random and infinite length condition. An important term k is used to describe block nature of RZL $\{m_0, m_1, m_2, ..., m_{k-1}\}$. We also introduce two kinds of coding used in Wong [2] and Zhang [3].

Table 3. Theoretical σ and λ under different RZL and different k

k	Zhang's RZL		Wong's RZL	
	Decrement rate of "1" σ=S_1/S_0= 2/(k+2)	Increment rate of length λ=L_1/L_0= $(2^{k+1}+2^k+1)/(2k+4)$,	Decrement rate of "1" σ=S_1/S_0= 2/k	Increment rate of length λ=L_1/L_0= $(2^k+1)/(2k)$
1	0.6667	1.1667	2.0000	1.5000
2	0.5000	1.6250	1.0000	1.2500
3	0.4000	2.5000	0.6667	1.5000
4	0.3333	4.0833	0.5000	2.1250
5	0.2857	6.9286	0.4000	3.3000

Notes: σ ---- Decrement rate of the number of "1" ,
S_0 is the number of binary "1" in original data, S_1 is the number of binary"1" in RZL code,
λ ---- Increment rate of length $\lambda = L_1/ L_0$,
L_0 is length of original data, L_1/ is length of RZL code,
k ---- Binary code length of digit number (cf. Table 1 and Table 2).

As shown in Table 3, there are only three RZL coding methods are feasible and practical for data hiding (in small frame). They are Zhang's k=1 (λ=1.1667, σ=0.6667); Zhang's k=2 (λ=1.6250, σ=0.5000); and Wong's k=3: (λ=1.5000, σ=0.6667). The RZL is not suitable in these cases, when the increment rate of length λ is too long in Wong's RZL k>3 coding and in Zhang's RZL k>2 coding, because the length increase is very sensitive for data hiding. Also as k<3 at Wong's RZL coding cannot be used, the decrement rate of "1" would not be less than one.

3 RZL Coding with Data Having Arbitrary Length

For realization of RZL, the original data should be complete, which means the length of last block of data should satisfy the definition. For example the length of the last block should be equal to k in Wong's RZL, or the length of the last block should be equal to k+1 in Zhang's RZL with prefix being "1". If the last block of original data cannot be cut off correctly, it will be hard to recover data from the RZL coding.

The method to solve the problem of incompletion of the last block is to add one or several "0" bits to form complete block and have some mark attached in RZL for lossless recover.

The experimental results show that the method to solve the problem of incompletion works. In Table 4, more examples are provided to explain the procedure of arbitrary length data cutting off for RZL.

There is a simplest example in Table 4: the original data m=1. The length doesn't satisfy the definition for m= {1}, which is too short and incomplete. It is hard to code incomplete block by RZL The attaching one 0 (underline) is used, so the last block becomes m= {10} (column 4 and line 2). It can satisfy the definition of coding of Zhang's RZL coding at k=1. We have two bit mark 01 (small frame) attached in RZL

c=101= {1[01]} (column 5 and line 2) which is used to show one bit 0 has been attached in original data for RZL lossless recover.

Table 4. The realization of two types of RZL

Data *m*	RZL type		To be completed, *m* is with attached "0" (underlined)	RZL code *c* with additional mark of the attached number of "0" by binary in frame
1	Zhang's		{10}	101={1}[01]
			{100}	110={1}[10]
	Wong's		{100}	0000110={00001}[10]
0101110	Zhang's		{0}{10}{11}{10}	00101100={00}{1}{01}{1}[00]
			{0}{101}{110}	00000100100={0000}{01}{001}[00]
	Wong's		{010}{111}{000}	00100000001110={001}{00000001}{1}[10]
010111	Zhang's		{0}{10}{11}{10}	00101101={00}{1}{01}{1}[01]
			{0}{101}{110}	00000100101={0000}{01}{001}[01]
	Wong's		{010}{111}	0010000000100={001}{00000001}[00]

Note: The additional mark of small frames in RZL code *c* in this table show the number by binary of underline "0" in completed *m*.

There is another example in Table 4: the original data m=010111(column 1 and line 4). The length is not satisfy the definition for m= {0}{101}{11}. The last block {11} is incomplete. It is hard to code incomplete block by RZL, even as all blocks prior to the last block haven't any problem. The attaching one 0 (underline) is used, so the last block becomes {110} or m= {0}{101}{110} (column 4 and line 9). It can be satisfy the definition of coding of Zhang's RZL coding at k=2 . We also have two bit mark [01] (small frame) attached in RZL c=00000100101={0000}{01}{001}[01] (column 5 and line 9) , which is used to show one bit 0 has been attached in original data for RZL lossless recover.

In Table 5, when using the method of RZL coding to code c under different length (200 bits, 2000 bits, 20000 bits), we use uniformly distributed binary random original data. We show the comparison between three-time each real random data (finite length) to theoretical one (infinite length, see in Table 4).

For example at Zhang's RZL k=1, the theoretical (infinite length) decrement rate of "1" is $\sigma=S_1/S_0=2/(k+2)=0.6667$ (column 4 and line 2~4),and in the real random data (finite length) σ is {0.7204，0.6792，0.6640} (column 3 and line 2~4). Also the theoretical (infinite length) increment rate of length is $\lambda=L_1/L_0=(2^{k+1}+2^k+1)/(2k+4)=1.167$ (column 5 and line 2~4), and in the real random data (finite length) λ is {1.225，1.152，1.157} (column 5 and line 2~4).

Table 5. Comparison of the real and the theoretical ones of Zhang's and Wang's RZL

Zhang's RZL	Times	L_0 (bit)	$\sigma=S_1/S_0$ (real, finite) (dB)	$\sigma=S_1/S_0$ $=2/(k+2)$ (theoretical, infinite) (dB)	$\lambda=L_1/L_0$ (real, finite) (dB)	$\lambda=L_1/L_0=$ $(2^{k+1}+2^k+1)/(2k+4)$ (theoretical, infinite) (dB)
k=1	1	200	0.6311	0.6667	1.225	1.167
		2000	0.6847	0.6667	1.152	1.167
		20000	0.6718	0.6667	1.157	1.167
	2	200	0.6509	0.6667	1.160	1.167
		2000	0.6517	0.6667	1.180	1.167
		20000	0.6632	0.6667	1.163	1.167
	3	200	0.6731	0.6667	1.130	1.167
		2000	0.6673	0.6667	1.156	1.167
		20000	0.6683	0.6667	1.165	1.167
k=2	1	200	0.5098	0.5	1.530	1.625
		2000	0.4970	0.5	1.639	1.625
		20000	0.4999	0.5	1.625	1.625
	2	200	0.5102	0.5	1.645	1.625
		2000	0.5062	0.5	1.668	1.625
		20000	0.5069	0.5	1.626	1.625
	3	200	0.5263	0.5	1.715	1.625
		2000	0.5029	0.5	1.544	1.625
		20000	0.4990	0.5	1.618	1.625
Wong's RZL	Times	L_0 (bit)	$\sigma=S_1/S_0$. (real, finite) (dB)	$\sigma=S_1/S_0=2/k$. (theoretical, infinite) (dB)	$\lambda=L_1/L_0$ (real, finite) (dB)	$\lambda=L_1/L_0$ $=(2^k+1)/(2k)$ (theoretical, infinite) (dB)
k=3	1	200	0.7204	0.6667	1.465	1.5
		2000	0.6792	0.6667	1.454	1.5
		20000	0.6640	0.6667	1.513	1.5
	2	200	0.7128	0.6667	1.470	1.5
		2000	0.6643	0.6667	1.521	1.5
		20000	0.6591	0.6667	1.512	1.5
	3	200	0.7128	0.6667	1.440	1.5
		200	0.6650	0.6667	1.497	1.5
		2000	0.6652	0.6667	1.499	1.5

In real experiment, the parameter value, σ and λ are much closer to that in the theoretical mostly. The longer the real original data length is, the closer the practical value will be to the theoretical value.

4 Histogram-pair Based Reversible Data Hiding Using RZL

In [4] a scheme to reversibly embed data into image's prediction-errors by using histogram-pair scheme is reported. Although the results outperform other reversible data hiding schemes, the coding to decrease the number of "1" for the data (such as the RZL) has not been utilized. We now use the RZL to improve the performance of data hiding [4] in this paper.

Fig. 2 shows the histogram-pair based reversible data hiding with the RZL coding, where x is original image, x' is the marked image.

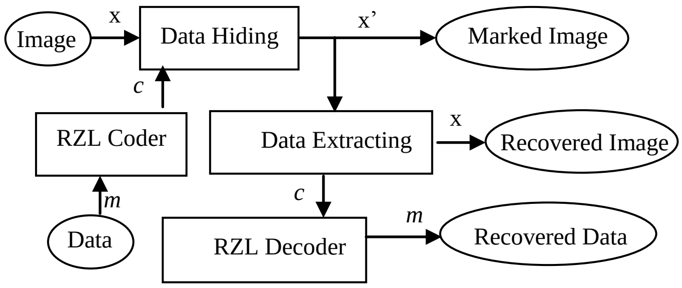

Fig. 2. Histogram-pair based reversible data hiding using RZL

The five BMP images, i.e., Woman (960×768), Lena (512×512), Barbara (512×512), Baboon (512×512) and Airplane (512×512) used in experiments are shown in Fig. 3. Tables 6 and 7 reported the PSNR under different payloads by using histogram-pair data hiding scheme with Non-RZL [4] and with RZL (Zhang *k*=1, Zhang *k*=2, Wong *k*=3) coding for the above-mentioned five BMP images. Note that Table 6 reports the image PSNRs after histogram-pair based reversible data hiding with small payload (0.0002 to 0.0007 bpp), while Table 7 reports the test results with median payload (from 0.01 to 0.2 bpp).

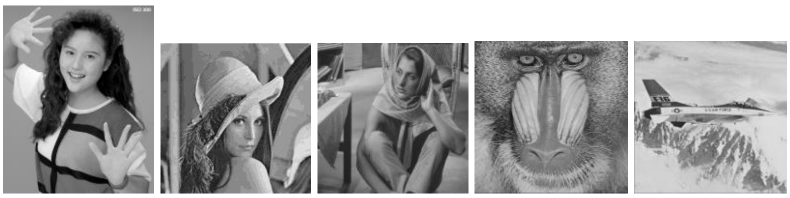

Fig. 3. BMP images from the left : Woman, Lena, Barbara, Baboon and Airplane

Table 6 is small payload under different payloads by histogram-pair data hiding scheme with Non-RZL [4] and three types of RZL coding for five BMP images. It is observed that for the small payload, the RZL coding does enhance the PSNR of the images with histogram-pair reversible data hiding. Note that, differently from other four images, the improvement of Baboon image is not obvious.

Table 6. Image PSNR after histogram-pair reversible data hiding by RZL with small payload

Payload (bpp)		0.0002	0.0003	0.0004	0.0005	0.0006	0.0007
Payload (bits) for images of 960×768		147	221	294	368	442	516
Payload (bits) for images of 512×512		52	78	104	131	157	183
Woman 960×768	Non-RZL	87.46	85.8	84.33	83.15	82.04	82.08
	Zhang's *k*=1	88.81	86.85	85.28	84.90	84.14	83.17
	Zhang's *k*=2	89.03	86.64	84.69	85.19	84.04	83.42
	Wong's *k*=3	88.36	86.09	84.50	84.61	83.32	82.71
Lena 512×512	Non-RZL	87.00	85.69	84.32	83.34	82.68	81.86
	Zhang's *k*=1	88.89	87.40	86.52	84.53	83.99	82.63
	Zhang's *k*=2	90.28	88.17	86.88	84.61	83.51	81.98
	Wong's *k*=3	89.31	88.00	85.78	84.39	82.82	81.67
Barbara 512×512	Non-RZL	85.60	83.93	82.63	81.75	81.60	80.95
	Zhang's *k*=1	87.00	84.99	84.68	83.23	82.32	81.31
	Zhang's *k*=2	87.69	84.68	84.12	82.97	81.79	81.05
	Wong's *k*=3	86.30	84.32	83.34	82.27	81.42	80.67
Baboon 512×512	Non-RZL	82.02	80.28	78.91	77.83	77.04	76.23
	Zhang's *k*=1	82.23	80.50	78.87	77.68	76.66	75.82
	Zhang's *k*=2	81.11	79.07	77.50	76.39	75.20	74.70
	Wong's *k*=3	79.94	78.37	77.00	75.87	75.10	75.44
Airplane 512×512	Non-RZL	86.19	85.07	84.12	83.4	82.73	82.32
	Zhang's *k*=1	87.69	86.52	85.78	84.83	84.32	83.74
	Zhang's *k*=2	88.70	87.69	86.88	85.78	84.83	84.39
	Wong's *k*=3	88.34	87.13	85.98	84.76	83.93	83.45

Note: The small frames in this table show that the PSNR of RZL is larger than PSNR of Non-RZL.

Table 7 reports the PSNR enhancement achieved with median payload by using histogram-pair data hiding scheme with Non-RZL [4] and three types of RZL coding for five BMP images. It is observed that the improvements achieved for median payload are not as large as the small payload in Table 6. Specially, it does not bring improvement for Baboon image at all.

Table 7. PSNR of histogram-pair scheme data hiding by RZL with median payload

Payload (bpp)		0.01	0.02	0.03	0.04	0.1	0.2
Payload (bit) for 960×768		7372	14745	22118	29491	73728	147456
Payload (bit) for 512×512		2621	5242	7864	10485	26214	52428
Woman 960×768	Non-RZL	69.22	64.67	63.15	62.05	58.38	54.35
	Zhang's *k*=1	70.9	65.15	63.63	62.52	58.70	54.45
	Zhang's *k*=2	70.87	64.77	63.24	62.16	55.75	51.63
	Wong's *k*=3	69.37	64.60	63.05	61.99	56.74	52.93
Lena 512×512	Non-RZL	67.08	63.78	61.79	60.38	55.44	50.87
	Zhang's *k*=1	67.31	63.82	61.82	60.26	55.11	50.10
	Zhang's *k*=2	66.20	62.70	60.51	58.97	52.13	47.57
	Wong's *k*=3	66.20	62.64	60.72	59.21	53.82	48.06
Barbara 512×512	Non-RZL	68.28	64.89	62.76	61.19	55.97	50.50
	Zhang's *k*=1	68.64	65.08	62.77	61.23	55.47	49.87
	Zhang's *k*=2	67.86	63.95	61.61	59.98	53.27	47.57
	Wong's *k*=3	67.73	63.92	61.71	60.14	53.54	47.98
Baboon 512×512	Non-RZL	63.51	59.57	57.29	55.46	47.65	41.60
	Zhang's *k*=1	63.04	58.98	56.49	54.31	46.55	40.61
	Zhang's *k*=2	61.52	55.88	53.42	51.51	43.92	37.54
	Wong's *k*=3	61.73	57.51	54.83	51.98	44.54	38.02
Airplane 512×512	Non-RZL	69.58	66.31	64.57	62.83	58.25	54.28
	Zhang's *k*=1	70.69	67.3	65.13	63.8	58.49	54.24
	Zhang's *k*=2	70.98	67.25	64.81	62.61	57.24	52.10
	Wong's *k*=3	70.42	66.79	64.58	62.87	57.33	52.30

Note: The small frames in this table indicate that the PSNR of RZL is larger than PSNR of Non-RZL.

Fig.'s 4~13 are performance curves for small and median payload under different payloads by histogram-pair reversible data hiding scheme with Non-RZL [4] and with three types of RZL coding for five BMP images. The PSNR gain of Fig.'s 4,6,8,10,12 on the left shown the small payload are higher than that in Fig.'s 5,7,9,11,13 on the right.

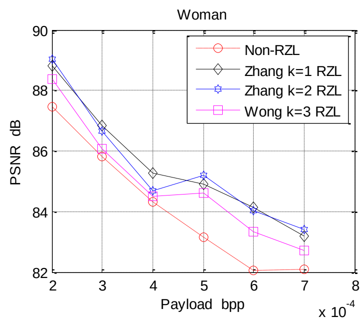

Fig. 4. Small payload into Woman by RZL

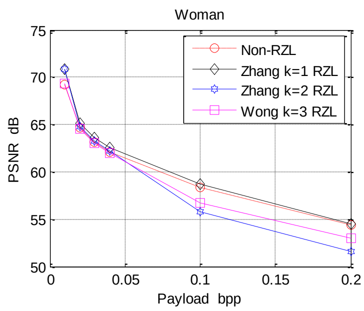

Fig. 5. Median payload into Woman by RZL

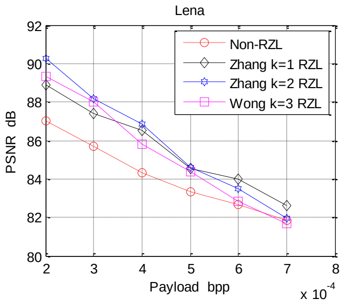

Fig. 6. Small payload into Lena by RZL

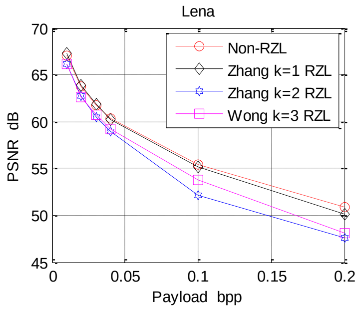

Fig. 7. Median payload into Lena by RZL

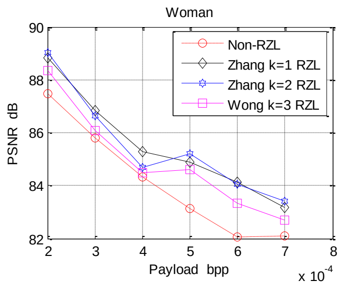

Fig. 8. Small payload into Barbara by RZL

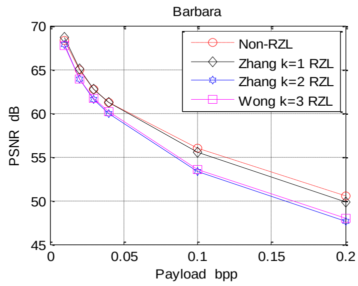

Fig. 9. Median payload into Barbara by RZL

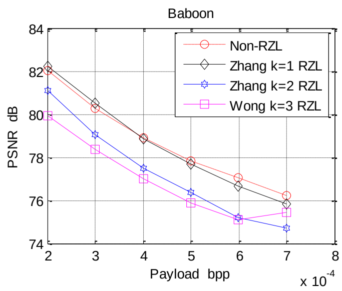

Fig. 10. Small payload into Baboon by RZL

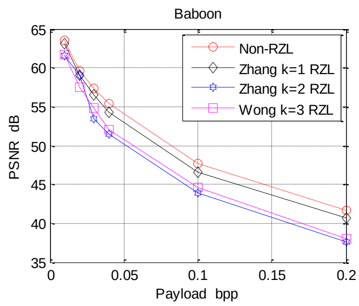

Fig. 11. Median payload into Baboon by RZL

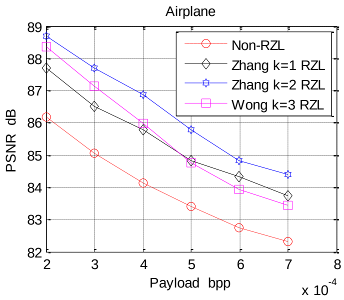

Fig.12. Small payload into Airplane by RZL

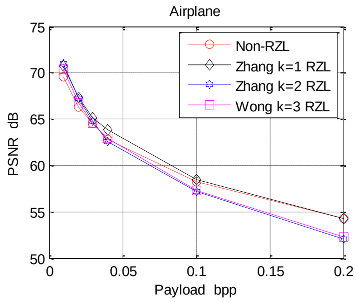

Fig. 13. Median payload into Airplane by RZL

Table 8 is average PSNR gain for different RZL to Non-RZL [4] under average payloads by histogram-pair data hiding scheme for five BMP images.

Table 8. Average PSNR gain for different RZL to Non-RZL

Image	Average of 6 small payload (0.0002 bpp~0.0007 bpp)			Average of 6 median payload (0.01 bpp~0.2 bpp)		
	Zhang's *k*=1 PSNR gain (dB)	Zhang's *k*=2 PSNR gain (dB)	Wong's *k*=3 PSNR gain (dB)	Zhang's *k*=1 PSNR gain (dB)	Zhang's *k*=2 PSNR gain (dB)	Wong's *k*=3 PSNR gain (dB)
Woman	1.3820	1.3580	0.7883	0.5780	-0.5533	-0.5233
Lena	1.5120	1.7570	1.1800	-0.1533	-1.8767	-1.4483
Barbara	1.1780	0.9733	0.3100	-0.0883	-1.5583	-1.4283
Baboon	-0.0920	-1.3900	-1.7650	-0.8500	-3.5483	-2.7450
Airplane	1.5080	2.4070	1.6270	0.6383	-0.1383	-0.2250
Grand average	1.0980	1.0210	0.4280	0.0249	-1.53498	-1.2740

Note: The small frames in Table 8 indicate that the PSNR of RZL is larger than PSNR of Non-RZL.

Table 9 is statistical analysis for different RZL and different average payloads and different images.

Table 9. Statistical analysis for different RZL

Item	Comparison
Different RZL and different average payloads	The average PSNR gain in small payload will be higher because PSNR is sensitive to decrement rate of the number of “1” σ (Zhang k=1：λ=1.1667，σ=0.6667; Zhang k=2：λ=1.6250，σ=0.5000; Wong k=3：λ=1.5000，σ=0.6667). For five images, the average PSNR gain of small payload is superior to that of median one obviously. Because the length of data by RZL is increasing rapidly on median size payloads. The average PSNR gain of Zhang’s RZL k=1 or k=2 is better than that of Wong’s RZL k=3.
Different images	Average PSNR gains are different for images. We only list PSNR gains the 6 payloads（0.0002 bpp~0.0007 bpp) for each image with 3 RZL: Zhang k=1, Zhang k=2, Wong k=3 respectively: Woman [1.382, 1.358, 0.7883], Lena [1.512, 1.757, 1.18], Barbara [1.178, 0.9733, 0.31], Baboon [-0.092, -1.39, -1.765], Airplane [1.508, 2.407, 1.627]. The Baboon is the worst one, maybe the texture of image is much rich and strong, and hence, the entropy of image is higher and the entropy is sensitive for data hiding by RZL.

5 Conclusions

1. There is feasibility and practicability for improvement of histogram-pair based image reversible data hiding scheme by using the RZL coding. This means that, after pre-processing to compress the random data to be embedded, the number of binary “1” will be decreased by RZL. The better image PSNR after reversible data hiding with the same embedding rate is realized for small payload in this paper.
2. The histogram-pair scheme of image reversible data hiding is just suitable to use RZL, because all zero distribution cover is given in histogram-pair scheme.
3. The method to solve the problem of incompletion of the last block is to add one or several “0” bits to form complete block and have some mark attached in RZL for lossless recover.
4. Average PSNR gain in small payload will be higher for most natural image. However, Baboon image is exceptional, i.e., not suitable for this method. Because the entropy of Baboon is higher, and the entropy is sensitive for data hiding by RZL.

5. Even the PSNR gain is higher only for small payload, the applications such as in authentication will not be serious problem sometimes. For example only 256 bit is needed in the Hash method SHA-256. When the image size larger than (512×512), for example (4000×3000), the small payload will be bigger and satisfied with many applications.
6. It is shown that the average PSNR of RZL is higher than Non-RZL in histogram-pair based image reversible data hiding. Zhang et al.'s RZL is better than Wong et al.'s RZL in our experimental work. With Zhang et al.'s RZL, the average PSNR gain is about 1 dB for five images.

References

1. Kalker, T., Willems, F. M.: Capacity Bounds and Code Constructions for Reversible Data-Hiding. Proc. of EI SPIE, Security and Watermarking of Multimedia Contents V, vol. 5020, Santa Clara, pp. 604-611, (2003)
2. Wong, K., Tanaka, K., Takagi, K., and Nakajima, Y.: Complete Video Quality-Preserving Data Hiding. IEEE Trans. on Circuits and Systems for Video Technology. 19(10):pp.1499-1512, (2009).
3. Zhang, W., Chen, B., and Yu, N., Capacity-Approaching Codes for Reversible Data Hiding" 13th Information Hiding Conference (IH 2011), May 18-20, 2011, Prague, Czech Republic
4. Xuan, G., Tong, T., Teng, J., Zhang, X., Yun Q. Shi, Optimal Histogram-pair and Prediction-error Based Image Reversible Data Hiding, 11th International Workshop on Digital-forensics and Watermarking (IWDW 2012), Oct 31-Nov. 3, 2012, Shanghai, China