A Novel Reversible Data Hiding Technique with High Capacity and Less Overhead Information

Hamida M. Almangush
Biomedical Computing and Engineering Technologies Applied Research Group, Centre for Advanced Computing and Telecommunication,Faculty of Information and Communication Technology Universiti Teknikal Malaysia Melaka

Mohd Khanapi Abd Ghani
Universiti Teknikal Malaysia Melaka

Ahmed Bashir Abugharsa
Universiti Teknikal Malaysia Melaka

ABSTRACT
In this paper, a new reversible image hiding scheme based on histogram shifting for grayscale images is proposed. As is known, the payload storage of histogram-based reversible data hiding is impacted by the overhead information of the pixel positions that have to be embedded in a cover image. To solve this problem, the cover image is divided into two parts, namely the Most Significant Part (MSP) and the Least Significant Part (LSP), secret data is hidden by shifting the histogram of the most significant part. To increase the payload of embedded data in a cover image, the proposed algorithm reduces the number of bits that represent the secret data without any corruption of that secret data. In addition, overflow and underflow problems are prevented by categorization of the histogram into three categories. According to the experimental results, the cover image is recovered correctly. A higher hiding capacity can be obtained and a good quality marked image is preserved when the proposed scheme is applied to hide the secret data by shifting the histogram of the most significant part instead of hiding by shifting the histogram of the whole cover image.

General Terms
Security.

Keywords
Reversible Data Hiding, Histogram Modification, Embedding capacity

1. INTRODUCTION
As long as there has been written communication, humans have had the desire to conceal their messages from the curious eyes of others. Moreover, the need for private and sufficiently secure communications in several applications such as e-banking, e-trading, mobile telephony, medical data interchange etc., is rapidly increasing [1]. With these forces driving it, research into information hiding has grown explosively. In medical imagining systems, in addition to perceptual transparency, it is desired to reverse the marked media back to the original cover media without any distortion after the hidden data is retrieved [2] because the distortion may cause the modified medical images to be unusable for further diagnosis [3].

The technical challenges of the reversible data embedding problem are found in increasing the capacity, maintaining the reversible characteristic and simultaneously decreasing the distortion of the original images [4].

A histogram-shifted-based lossless data hiding algorithm was proposed by Ni et al. They utilized the zero (or the minimum) points of the histogram to embed data bits in a cover medium. The resulting peak signal-to-noise ratio (PSNR) was about 48.20 dB, the hiding capacity was not sufficient although their method caused only a slight distortion with low complexity [3]. T.C. Thanujathe et al. proposed reversible data hiding using an increased peak histogram. Their proposed algorithm made use of the assumption that most of the gray values around the peak have a relatively large number of pixels [5], thus the embedding capacity can be increased by increasing the number of pixels at the peak. Their proposed algorithm achieved good imperceptibility. However the weakness in this algorithm was the large amount of overhead information – in the form of the positions of the pixels around the peak – that had to be saved in order to recover the original image.

In this paper, we show that by applying reversible data hiding based on a shifted-histogram of an image, not only can the watermarked image quality be improved, but more importantly, the data hiding payload can be significantly increased. The proposed algorithm divides the cover image into two parts in the form of a Most Significant Part (MSP) and a Least Significant Part (LSP). The secret message is hidden by shifting the histogram of the most significant part, and therefore the distortion in the cover image is low. We also adopt a histogram shifting technique to prevent overflow and underflow problems. In histogram-based algorithms, the maximum payload data is the number of peak points. So the orientation of the conventional histogram-shifting based schemes is to find more peak point to increase the embedded data. We propose to minimize the representation of the secret message. The proposed algorithm minimizes the secret message size by 36.84 % of its size.

The rest of the paper is organized as follows. The characteristics of the proposed algorithm and its details are described in Section 2. Experimental results are presented in Section 3, and Conclusions are drawn in Section 4.

2. METHODS
In general, payload capacity and marked image quality are the important criteria in the evaluation of reversible data embedding. Payload capacity refers to how much data can be embedded in an image. In practice, a high payload capacity and low distortion are conflicting requirements; the larger the capacity created by a reversible embedding technique, the higher the distortion introduced by the embedding [6].

To increase the payload capacity in a cover image we propose to reduce the overhead information that is needed at the receiver side to reconstruct the original image from the marked image and also reduce the number of bits that represent the secret data by using the trinary number system instead of the binary number system. Trinary is the base-3 numeral system which the three digits 0, 1, and 2 are all whole numbers. Analogous to a bit, a trinary digit is a trit [7]. One trit contains about 1.5833 bits of information. the trinary format for symbols encoding is called "tryte" (analog of binary byte) consisting of 6 trits (~9.5 bits) ,so each 19 bits represented in the binary code can be represented in only 12 trits by using the trinary code, thus the secret data size will be reduced by 36.84 % of its size. For example,

Binary code : 111111111111111111

Dismal code : 524287

Trinary code : 222122012001

The proposed method consists of preprocessing, embedding, extraction, and recovery algorithms.

2.1 The Pre-processing Algorithm

The preprocessing step is performed on the original image.

2.1.1 The Most Frequent Colour

The number of pixel values with a gray-level of between 0 and 255 is calculated and denoted by h(0),…h(255). The top point value is searched from {h(0),…h(255)}, that is, h(t) = max {h(0),…h(255)}. The T value is saved as overhead information O1.

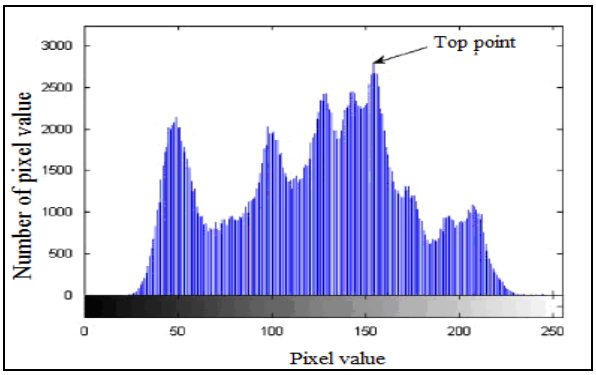

Figure 1: Histogram of an image with a peak point of 154

2.1.2 The Most Significant Part (MSP)

The Most Significant Part (MSP) is determined which includes all pixels with a value of the top point T. The most significant part is denoted by X1, X2, Y1 and Y2 that will be saved as overhead information O2.

Shifting the most significant part histogram instead of shifting the whole cover image histogram decreases the overhead information that is required in order to be able to reverse the marked image to retrieve the original image. Figure 2 presents the most significant and the least significant parts of a cover image.

Assuming X is the least frequent colour in the cover image histogram that is *h* (x)=min { h(0),.., h(255)}, then

$$h(x) = h\text{MSP(x)} + h\text{LSP(x)} \quad (1)$$

$$hMSP(x) = h(x) - hLSP(x) \quad (2)$$

In the case when *h*MSP(x) is less than *h* (x), the overhead information will be decreased.

Assume M is the least frequent colour in the histogram of the most significant part, that is *h*MSP(m)=min {hMSP(0),.., hMSP (255)}. In the case when *h*MSP (m) is less than *h* (x) the overhead information will be decreased.

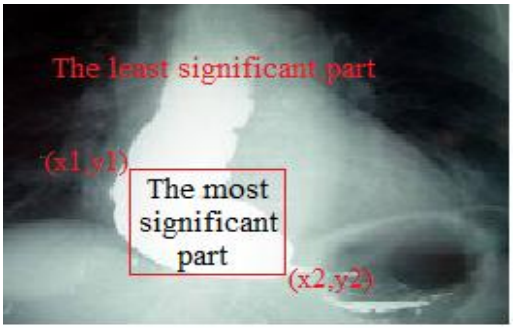

Figure 2: The most significant part and the least significant part of a cover image

2.1.3 The Least Frequent Colour

The least frequent colour value M1 is searched for in the most significant part (MSP) of the cover image, that is h(m1)=min {h(0),…., h(255)}. The M1 value and the location information of the pixels with a value of M1 are saved as overhead information O3. Pixels that are within the most significant part of the cover image and with values equal to M1 are changed by the addition or subtraction of 1 as follows:

Case 1: the T value is bigger than M1 value. If M1 is T-1, no pixels are shifted, otherwise all Pixel values from T-1 to M1+1 shift to M1 by 1.

Case 2: the T value is less than M1 value. If M1 is T+1, no pixels are shifted, otherwise pixel values from T+1 to M1-1 shift to M1 by 1.

2.1.4 The Second Least Frequent Colour

The least frequent colour value M2 is searched for in the most significant part (MSP) of the cover image, that is h(m2)=min{h(0),….,.h(255)}. The M2 value and the location information of pixels with a value of M2 are saved as overhead information O4. Pixels that are within the most significant part of the cover image and with values equal to M2 are changed by the addition or subtraction of 1 as follows:

Case 1: the T value is bigger than the M1 and M2 values. If M2 is T-2, no pixels are shifted, otherwise all pixel values from T-2 to M2+1 shift to M2 by 1.

Case 2: the T value is bigger than the M2 value and less than the M1 value. If M2 is T-1, no pixels are shifted, otherwise all pixel values from T-1 to M2+1 shift to M2 by 1.

Case 3: the T value is less than the M1 and M2 values. If M2 is T+2, no pixels are shifted, otherwise all pixel values from T+2 to M2-1 shift to M2 by 1.

Case 4: the T value is less than the M2 value and bigger than the M1value. If M2 is T+1, no pixels are shifted, otherwise all pixel values from T+1 to M2-1 shift to M2 by 1.

The values of M1, M2 and T identify the category of the most significant part of the cover image as follows:

- Category 1: the T value is between the M1 and M2 values.
- Category 2: the T value is bigger than the M1and M2 values.

- Category 3: the T value is less than the M1 and M2 values.

Figures 3-5 show the three histograms categories and each category after shifting to create two gaps that will be modified by the embedding algorithm to represent the secret data.

The distortion in the cover image is dependent on the numbers of modified pixels that are between the peak point and the two least frequent colours within the most significant part of the cover image.

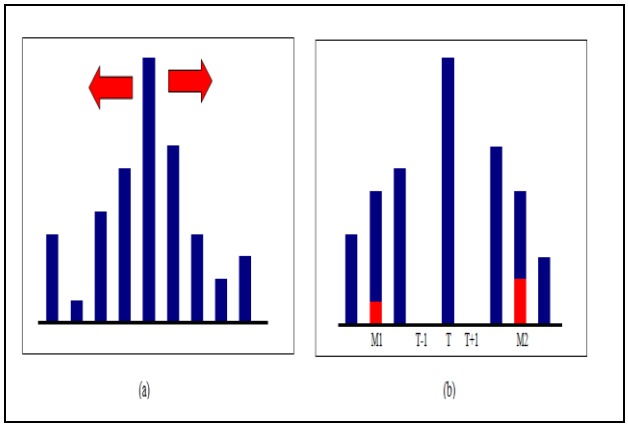

Figure 3: The first histogram category

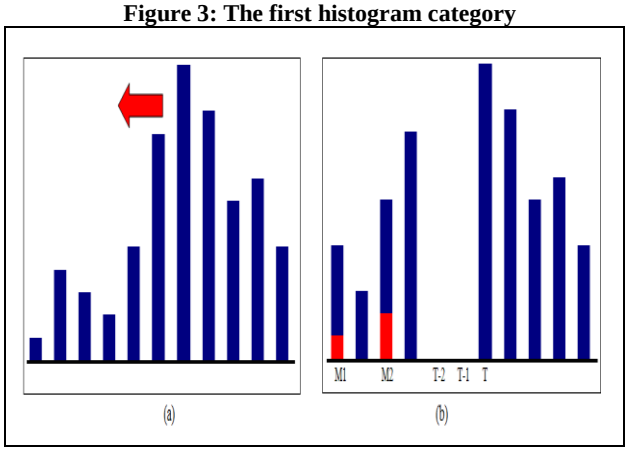

Figure 4: The second histogram category

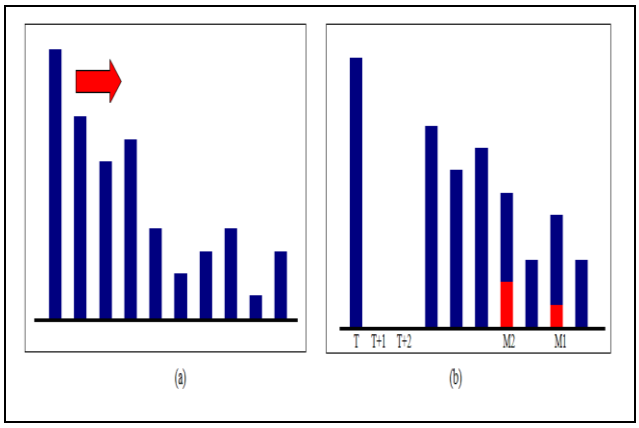

Figure 5: The third histogram category

2.2 The Embedding Algorithm

Secret data to be embedded along with the overhead information and the length of the secret data are represented by the ternary number system, that is, the base-3 number system. Thus it is represented by the set of {0,1,2}. The most significant part within a cover image that is indicated by X1, X2, Y1 and Y2 are scanned, and the pixels that have a value equal to the top point T in the cover image histogram are changed to represent the secret data.

The secret data d is embedded in the pixels P, and the modified pixels are called P'. Data is embedded according to the histogram category as follows:

Category 1: $$p' = \begin{cases} p - d & if\ d \leq 1 \\ p + d - 1 & otherwise \end{cases} \quad (3)$$

Category 2: $$p' = p - d \quad (4)$$

Category 3: $$p' = p + d \quad (5)$$

2.3 The Extraction Algorithm

The hidden data is extracted from pixels that are within the part of the marked image that is denoted in the overhead information O2. From the M1, M2 and T values that are stored within the overhead information O1 and O3, the category of the most significant part histogram from the original cover image is determined. The hidden data is extracted according to the histogram category as follows:

Category 1: T value is between the M1 and M2 values. Data is extracted from pixels P' that have values ε {T,T+1,T-1}.

$$d = \begin{cases} T - p' & if\ p' \in \{T, T-1\} \\ T - p' + 3 & otherwise \end{cases} \quad (6)$$

Category 2: T value is bigger than the M1 and M2 values. Data is extracted from pixels P' that have values ε {T,T-1,T-2}.

$$d = T - p' \quad (7)$$

Category 3: T value is less than the M1 and M2 values. Data is extracted from the pixels P' that have values ε{T,T+1,T+2}.

$$d = p' - T \quad (8)$$

The extracted data is represented by the trinary code {0,1,2}. To convert it to the binary code, each 12 trits within the extracted data are changed to 19 bits in the binary code{0,1}, thus the data obtained is same as the secret data.

2.4 The Recovery Algorithm

The recovery algorithm is the inverse of the preprocessing algorithm. According to the histogram category obtained in the extraction algorithm, the pixel values that lie within the part of the marked image as determined by the overhead information X1, X2, Y1 and Y2 are re-shifted to their original values.

Figures 6(a)-(c) show an example of recovery the original image from the marked image. The recovery algorithm is in three steps as follows:

- Determine the ranges of the modified pixels as below:
  Category 1: if M2>T,R1={ T+1,...., M2 },
  R2={M1,....,T- 1};otherwise
  R1={ M2 ,...,T-1}, R2={T+1,.. , M1}.
  Category 2: R1={M2 ,….,T- 2}, R2={M1 ,...,T -1}.
  Category 3: R1={ T+2 ,…, M2}, R2={T+1,....,M1}.

- Re-shift the modified pixels that are within R1, except the pixels that have positions within O4.

$$p = p' + 1,\ p' \in R1 \quad (9)$$

- Re-shift the modified pixels that are within R2, except the pixels that have positions within O3.

$$p = p' + 1,\ p' \in R2 \quad (10)$$

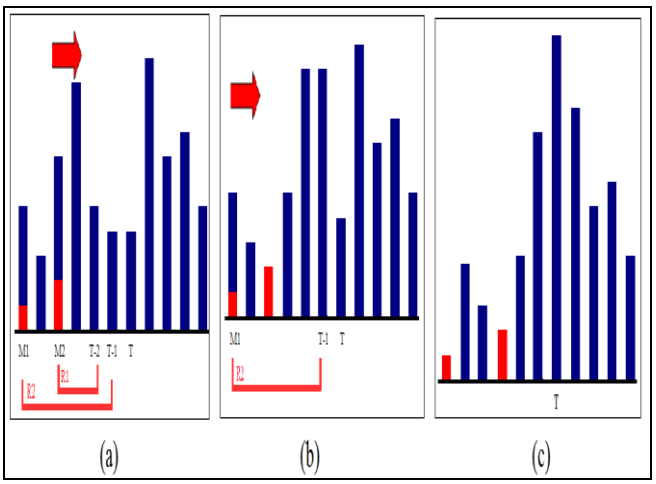

Figure 6: Recovery of the original image from the marked image. (a) Ranges of the modified pixels. (b) Image histogram after re-shift R1. (c) Image histogram after re-shift R2

3. EXPERIMENTAL RESULTS

We have successfully applied our proposed algorithm to commonly used grayscale images such as Lena and Baboon, etc.,. There is no salt-and pepper noise in all of tests since the proposed algorithm does not use modulo-256 addition. The proposed algorithm also solves the underflow and overflow problems. The two important criteria, the amount of payload (data to be hidden) and the degree to which cover media becomes distorted are used to evaluate the proposed technique [8]. The payload is computed by the following equation:

$$p = 1.583 \times h(\mathrm{T}) \quad (11)$$

where *h*(T) is the number of pixels, which is associated with the peak point [9]. The embedding capacity of the pixels at the peak point is 1.583 bpp (bits per pixel).

In our experiments, the quality of the marked image is measured by the peak signal-to-noise ratio (PSNR) which is the most popular criterion to measure the distortion between the cover image and marked image.

The PSNR is often expressed as a figure on a logarithmic scale in decibels (dB). PSNR values falling below 30dB indicate a fairly low quality, i.e., distortion caused by embedding can be obvious. A high quality marked image should strive for a PSNR value of 40dB and above [10]. A higher PSNR value can guarantee less distortion caused in the cover image[6].

$$\mathrm{PSNR} = 10\mathrm{Log}_{10}\left(\frac{255^2}{\mathrm{MSE}}\right) \quad (12)$$

where MSE denotes the Mean Square Error which is given as:

$$\mathrm{MSE} = \frac{1}{\mathrm{MN}}\sum_{x=1}^{M}\sum_{y=1}^{N}\left(\mathrm{S}_{xy} - \mathrm{C}_{xy}\right)^2 \quad (13)$$

where x and y are the image coordinates, M and N are the dimensions of the image, Sxy is the generated marked image and Cxy is the cover image [11].

We applied the proposed algorithm to embed the maximum payload within the cover image and within the most significant part of the cover image. The test results show that the overhead information of the pixel positions is decreased in all the tests when the secret data is embedded within the most significant part of the cover image. Table I summarizes the tests results.

Figure 7(a)-(c) and 7(d)-(f) shows the original images and the marked images. It can be seen that the marked images are very similar to the original images. Figure 8(a)(c)(e) shows the modified pixels when the secret data is hidden by shifting the histograms of the cover images. Figure 8(b)(d)(f) shows the modified pixels when the secret data is hidden by shifting the histograms of the most significant parts (MSP) of the cover images.

A comparison with other related methods, using similar images where possible, is tabulated in Table II. The table shows that the proposed algorithm produces the least visual distortion on the cover images (higher PSNR value).

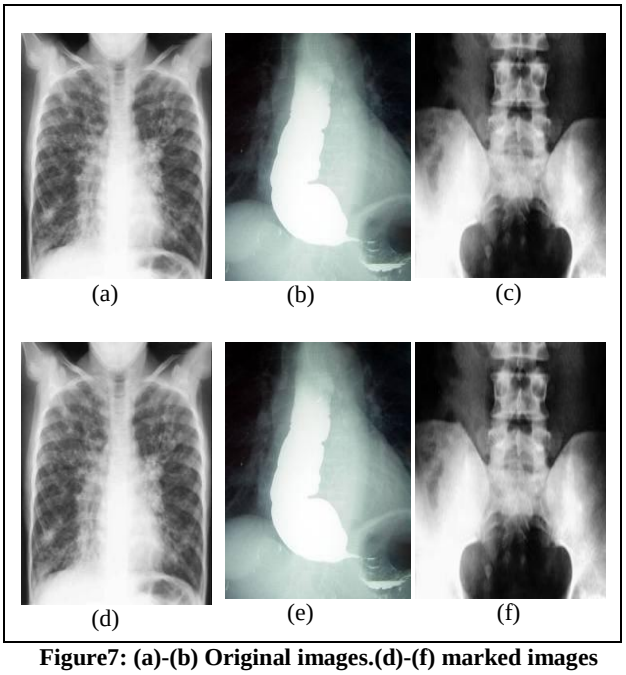

Figure7: (a)-(b) Original images.(d)-(f) marked images

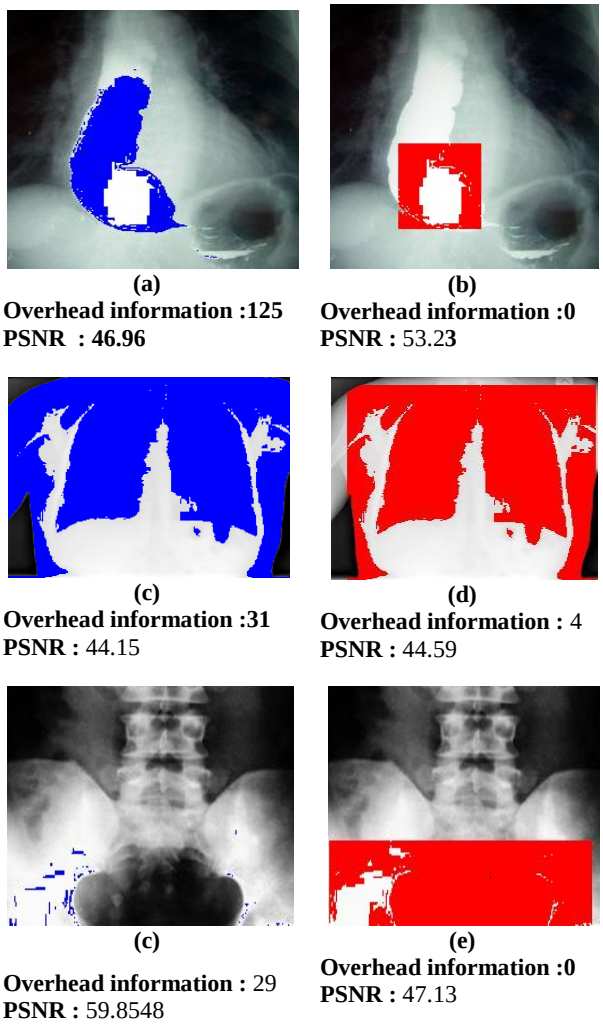

Figure 8:(a)(c)(e) the modified pixels in cover images, (b)(d)(f) the modified pixels in the most significant parts of the cover images

Table I. Overhead information and distortion comparison

Image	The whole cover image		The most significant part	
	h(M1)+h(M2) (Pixel)	PSNR (db)	*h*(M1)+h(M2) (Pixel)	PSNR (db)
1	90	42.27	0	49.05
2	39	44.08	28	47.30
3	35	42.25	0	44.36
4	34	42.43	0	51.10
5	32	48.91	27	52.11
6	20	43.64	17	48.01
7	17	44.83	0	46.46
8	0	48.15	0	50.61
9	0	48.17	0	55.41
10	0	42.21	0	49.22
11	0	47.90	0	50.25
12	0	42.64	0	49.85

Table II. Pure payload and distortion comparisons with other methods

Method	Lena (512 x 512 x 8)		Baboon (512 x 512 x 8)	
	Payload (bits)	PSNR (db)	payload (bits)	PSNR (db)
Goljan et al.	24,108	39	2,905	39
Vleescchouwer et al.	1,024	30	1,024	29
Xuan et al.	85,507	36.6	14,916	32.8
Celik et al.	74,600	38	15,176	38
Zhicheng et al.	5,460	48.2	5,421	48.2
Proposed	12,613	52.6	7,041	48.5

4. CONCLUSION

In this paper, we propose a novel reversible data hiding scheme based on a histogram technique with a very high capacity and low distortion in the cover image. Experimental results show the performance of our method reduces the bits that represent the secret data without any corruption and reduces the overhead information that affects the payload storage by shifting the histogram of the most significant part (MSP) of a cover image instead of shifting the histogram of the whole cover image. The marked image appears the same as the cover image, and it is reversible to retrieve the original cover image. The distortion in the cover image is dependent on the number of modified pixels that are between the most frequent colour and the two least frequent colours within the most significant part of the cover image.

5. ACKNOWLEDGMENTS

This paper is part of PhD work in the Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka (UTeM).

6. REFERENCES

[1] T. Ang and B. Delina, "WhiteSteg: a new scheme in information hiding using text steganography," WSEAS Transactions on Computers, vol. 7, pp. 735-745, 2008.

[2] Z. Ni, et al., "Reversible data hiding," Circuits and Systems for Video Technology, IEEE Transactions on, vol. 16, pp. 354-362, 2006.

[3] S. C. Huang and M. S. Lin, "A High-capacity Reversible Data-hiding Scheme for Medical Images," Journal of Medical and Biological Engineering, vol. 30, pp. 289-295, 2010.

[4] J. Y. Hsiao, et al., "Block-based reversible data embedding," Signal Processing, vol. 89, pp. 556-569, 2009.

[5] T. Thanuja, et al., "Reversible Data Hiding using Increased Peak Histogram," 2008, pp. 44-47.

[6] C. Chang, et al., "Reversible data-embedding scheme using differences between original and predicted pixel values," Information Security, IET, vol. 2, pp. 35-46, 2008.

[7] N. Brousentsov, et al., "Development of ternary computers at Moscow State University," Russian Virtual Computer Museum. Ed. Eduard Proydakov, vol. 11, 2002.

[8] W. Chung-Ming and W. Peng-Cheng, "Data hiding approach for point-sampled geometry," IEICE Transactions on Communi-cations, pp. 190-194, 2005.

[9] R. Ramaswamy and V. Arumugam, "Lossless Data Hiding Based on Histogram Modification," The International Arab Journal of Information Technology (IAJIT), 2010.

[10] Adnan Mohsin Abdulazeez Brifcani and W. M. A. Brifcani, "Stego-Based-Crypto Technique for High Security Applications," International Journal of Computer Theory and Engineering, vol. Vol.2, pp. 835-841, 2010.

[11] Y. K. Jain, "A Novel Image Steganography Method With Adaptive Number of Least Significant Bits Modification Based on Private Stego-Keys," International Journal of Computer Science and Security (IJCSS), vol. 4, p. 40, 2010.