MIS Review Vol. 18, No. 1, September (2012), pp. 57-69

Data Hiding Technique Based on LSB Matching towards High Imperceptibility

Marghny H. Mohamed[1], Naziha M. Al-Aidroos[2], Mohamed A. Bamatraf[3]

[1]*Department of Computer Science, Asyut University*

[2]*Department of Computer Engineering, Hadhramout University of Science and Technology*

[3]*Department of Computer Science, Hadhramout University of Science and Technology*

ABSTRACT: *Steganography is the art and science of hiding data. The secret message is hidden in such a way that the observer cannot detect any changes in the original image. In this paper, we propose an efficient steganographic scheme which provides high capacity of secret data as well as imperceptibility of stego image. Using fixed number (i.e., max) as the upper limit criteria for embedding, the target pixels selected for embedding are based on the number of bits which matches between the secret data bits and the cover pixel bits. As an indicator to determine which pixel is used for embedding, the first bit is reversed (negated). The experimental results over greyscale images showed, the ability of embedding high data capacity with preserving stego image quality. Efficiency of the model is evaluated using two metrics, the Peak-Signal-to-Noise Ratio (PSNR) value as one of the evaluation metrics, and the visual effects over the cover image as the second. Results are drawn and compared with one of the most common techniques (Classic least-significant bit, LSB) and accordingly showed significant advancement.*

KEYWORDS: *Steganography, Data Hiding, Data Security, LSB (least-significant bit) Substitution, LSB Matching.*

1. Introduction

The rapid development of the internet and the digital information revolution caused significant changes in the global society, ranging from the influence on the worldwide economy to the way people nowadays communicate. Broadband communication networks and multimedia data available in a digital format (images, audio, video) opened many challenges and opportunities for innovation.

In order to keep the security and privacy of our data from unauthorized access, variety of techniques have been proposed in the field of data security. Cryptography and Steganography are two main methods in this field.

Cryptography is the process of transforming a secret data into a form that is unreadable, except by someone who has a proper key. For any unauthorized user who does not have a key, the ciphertext will look like nothing but streams of meaningless code. Although cryptography is a good way to secure data, it still has some weaknesses. The appearance of ciphertexts would give unauthorized user an impulse to recover them. Moreover, the unauthorized users might even simply destroy the ciphertext out of range when they have trouble recovering them so that the legal receivers cannot get the data in time.

Steganography is a branch of information hiding. It embeds the secret message in the cover media (e.g., image, audio, video, etc.) to hide the existence of the message. The word Steganography comes from the Greek words steganos and graphia, which together means "hiding writing."

Confidentiality is at the heart of what steganography does. Steganography, though, accomplishes confidentiality in a slightly different manner than cryptography. With cryptography, an unauthorized person can see the information but cannot access it. Because he or she can tell that there is information being protected, the unauthorized person may try to break the encryption. With steganography, because the data is hidden, any unauthorized party does not even know there is sensitive data there. From a confidentiality standpoint, steganography keeps the information protected at a higher level. In brief, we can say that while cryptography is about protecting the content of messages, steganography is about concealing their very existence.

The digital steganography process has three basic components:

(1) The data to be hidden (secret data).

(2) The cover file (cover-carrier), in which the secret data are to be embedded.

(3) The resulting stego-file (stego-carrier).

For the past decade, many steganographic techniques for still images have been presented. A simple and well-known approach is directly hiding secret data into the least-significant bit (LSB) of each pixel in an image. Then, based on the LSB technique, a genetic algorithm of optimal LSB substitution is available to get better stego-image quality than the simple LSB method (Wang, Lin, & Lin, 2001). In addition, Chang, Hsiao, and Chan (2003) proposed a fast and efficient optimal LSB method based on the dynamic programming strategy that improves the computation time of Wang et al.'s scheme (Wang et al., 2001).

Thien and Lin (2003) also presented a simple LSB scheme based on the modulus function to improve the stego-image quality. Besides, A novel simple LSB technique based on optimal pixel adjustment was presented to achieve the goal of improving the stego-image quality (Chan & Cheng, 2004).

Liu et al. (2004) presented a novel bit plane-wise data hiding scheme using variable-depth LSB substitution and employed post-processing to eliminate the resulting noticeable artifacts.

In order to gain a higher payload than when the 4-LSBs method is used, Wang (2005) has proposed two new schemes based on the modulo operator.

In order to enhance the security, on the other hand, Lou and Liu (2002) proposed a LSB-based steganographic method that can resist the common-cover-carrier attack by embedding variable-size secret data and redundant Gaussion noise. Lin and Tsai (2004) have proposed a new approach that integrates the concept of secret image sharing and steganographic techniques with the additional capability of image authentication.

The LSB-based methods mentioned above, directly embed the secret data into the spatial domain in an unreasonable way without taking into consideration the difference in hiding capacity between edge and smooth areas. In general, the alteration tolerance of an edge area is higher than that of a smooth area. That is to say, an edge area can conceal more secret data than a smooth area. With this concept in mind, Wu and Tsai (2003) presented steganographic scheme that offers high imperceptibility to the stego-image by selecting two consecutive pixels as the object of embedding. The payload of Wu and Tsai's scheme is determined by the difference value between the pixels. In a paper by Zhang and Wang (2004) Pixel Value Differencing (PVD) was successfully attacked. This was done through an analysis of the histogram of the stego image. Wu et al. (2005) proposed a method based on PVD which tries to increase the embedding capacity of PVD. They used LSB embedding for smooth regions and PVD embedding for edged areas. Wang et al. (2008) improved the stego-image quality by adjusting the remainder of the two consecutive pixels instead of the difference value, by using the modulus operation. Besides that, this method solved the falling-off-boundary problem also.

An Adaptive data hiding method was proposed by Yang et al. (2008). In this method pixels located in the edge areas are embedded by a k-bit LSB substitution method with a larger value than that of the pixels located in smooth areas.

An effective steganographic scheme has to be implemented that thwarts the attacker from extracting the secret information during transmission and reception (Chen, Chang, & Le, 2010).

Padmaa and Venkataramani (2010) proposed a methodology which enhances the Chang and Tseng (2004) technique by increasing the embedding capacity and improves the stego image quality using Thein and Lin (2003) algorithm by adapting Zig-Zag traversing scheme (ZZTS).

In this paper, 8-bit grayscale images are selected as the cover images. For data hiding methods, the image quality refers to the quality of the stego images. The proposed method is based on LSB substitution method and on finding a matching between the secret data bits and the cover pixels bits. To prevent illicit access of the data and to obtain better stego image quality, our method is presented.

The rest of this paper is organized as follows. Section 2 briefly describes the simple LSB substitution. In Section 3, the proposed method is presented. Experimental results with a brief discussion are given in Section 4. Finally conclusions are presented in Section 5.

2. LSB steganography

LSB insertion is a common and simple approach to embed information in a cover file: it overwrites the LSB of a pixel with an *M*'s bit. Unfortunately, modifying the cover image changes its statistical properties, so eavesdroppers can detect the distortions in the resulting stego image's statistical properties. The general operation of data hiding by using a simple LSB substitution method is described in this section.

Many steganogrphic methods embed a large amount of the secret information in the first *k* LSB's of the cover image pixels. Because of the imperfect sensibility of the human visual system, the existence of the embedded secret information can be imperceptible.

A digital image *I* can be represented by a two dimensional, with size (N = *width* × *height*). Each image is composed of finite elements each of which has a definite location and amplitude. These elements are referred to as image pixels, $I = \{P_1, \ldots, P_N\}$, where every pixel P_i in the greyscale image consists of 8 bits: $|P_i| = 8$, $P_i = \{b_1, ..., b_8\}$, where: $b_i \in \{1, 0\}$.

LSB with *k* embedding factor $1 \leq k \leq 8$, for every P_i targeted for embedding data bits replaces the set of bits: $T = t_1, ..., t_k$, keeping the rest of bits $\overline{T}$ = (*i.e.*, t_{k+1} ..t_8) without effect.

The generated set of pixels $\{P'_1, ...,P'_N\}$ represents as the stego image I', where: $P'_i = \{b'_1, ...b'_8\}$, $b'_j \in \{1, 0\}$.

The following example describes how the LSB embedding happens, suppose we have the following pixels:

P_1 = [11001011], … P_2 = [00011010], … P_3 = [01001100].

And the bits want to embed it in the LSBs positions are M = [010], the resulted pixels after embedding will be:

$P_1 = [1100101\underline{0}]$, … $P_2 = [0001101\underline{1}]$, … $P_3 = [0100110\underline{0}]$.

The quality of the stego image produced by simple LSB substitution may not be acceptable. It means that the method degrades the image quality and probably attracts unauthorized attention. Once he/she notices the stego image, secret message can be easily extracted by simple LSB analysis. The proposed method can solve these problems, even though it is based on classic LSB, yet it is totally different as the embedding process is not a uniform in terms of what pixels are selected for embedding in the cover image. This increases the complexity of the hidden data extraction in one hand and preserves the image quality in the other hand, by minimizing the number of the modified LSB bits.

3. The proposed scheme

Any image I consists of set of pixels:

$$I = \{P_1, \ldots, P_N\}, |P_i| = 8 \text{ bits}, \quad P_i = \{b_1, .., b_8\}, b_j \in \{1, 0\}. \tag{1}$$

The image size is computed as:

$$N = W \times H \tag{2}$$

Where W, H is the image width and height respectively. Suppose M is the secret data bits, with length n,

$$M = \{m_1, m_2, ..., m_n\}, \text{ where } m_i \in \{1, 0\}. \tag{3}$$

Like other data hiding schemes, the proposed scheme consists of two procedures; the embedding procedure and the extracting procedure. In this section the procedures steps are described in detail.

3.1 The embedding procedure

The data embedding process' steps are explained below:

Step 1: To embed secret data bits, firstly, select a number say *max*, where ($1 \leq max \leq 8$) to define the similarity threshold to be targeted for embedding.

Step 2: Scan pixel bits starting from 2nd LSB bit, to check if this pixel is valid for embedding or not based on embedding criteria (*i.e.*, *matching*) as follows:

$\forall$ cover image pixels P_i , where ($1 \leq i \leq N$),

(1) $\forall P_i\,[j] \in P_i$, where ($2 \leq j \leq 8$), if $P_i\,[j] = M_k$, so go to the next bit by increase the bit's counter (l) by 1, this step will be repeated until unmatched bit is found or until $l = max$ (*i.e.*, *max*: number of matches achieved between pixel bits and secret data bits).

(2) If $l = max$, meaning the embedding is done, so go to the next pixel.

Step 3: If the scanned pixel is valid for embedding then reverse the 1st LSB bit, else all pixel bits will remain unchanged.

$$\begin{aligned} &\mathit{if}\,(P_i\,[1] == 0) \\ &\quad P_i\,[1] = 1, \\ &\mathit{else}\ P_i\,[1] = 0 \end{aligned} \tag{4}$$

Step 4: After embedding all the secret data bits, the stego image is generating. The block diagram of the entire embedding procedure is represented in Figure 1.

3.2 The extracting procedure

This procedure is done in the other communication side to extract the hidden secret data bits, as following:

Step 1: Using the original cover image, the comparing between cover (I) and stego (I')

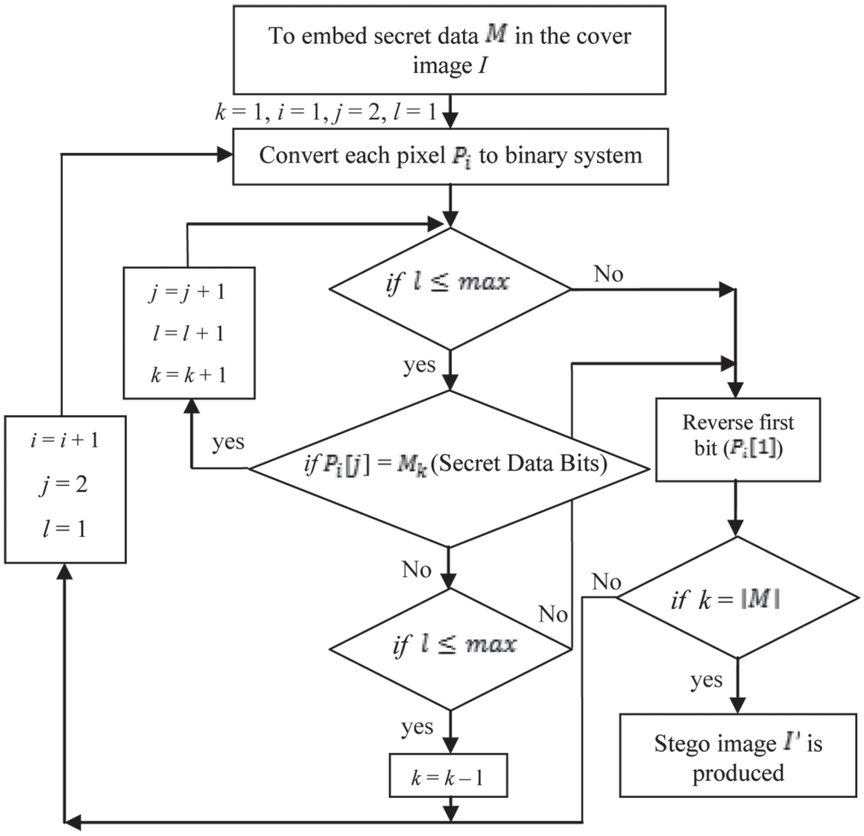

To embed secret data M in the cover image I
$k = 1, i = 1, j = 2, l = 1$
Convert each pixel P_i to binary system
$if\ l \le max$
No
yes
$j = j + 1$
$l = l + 1$
$k = k + 1$
yes
$i = i + 1$
$j = 2$
$l = 1$
$if\ P_i[j] = M_k$ (Secret Data Bits)
Reverse first bit ($P_i[1]$)
No
No
No
$if\ k = |M|$
$if\ l \le max$
yes
yes
$k = k - 1$
Stego image I' is produced

Figure 1 A Block Diagram of Embedding Steps

image pixels is done. Any pixel's value in the stego image differ from its original value, meaning it has an embedded bits.

```
    for ($i = 1$ to $N$)
      if ($P_i != P'_i$)
  $P'_i$ has an embedded bits,                    (5)
    else go to next pixel
```

Where P_i, P'_i is the cover, stego pixel respectively.

Step 2: If the embedded pixels are determined, the extracting is done by extracting the *max* bits from it.

```
    $j = 2, k = 1,$
  for ($l = 1$ to $max$)
    $M'_k = P'_i[j],$                             (6)
  $k = k + 1, j = j + 1.$
```

Where M'_k is the extracted secret message, k is the index of it, $P'_i\ [j]$ is the jth bit of the stego image pixel P'_i , and l is the counter of bits to be extracted.

Step 3: At this stage, the retrieving algorithm finishes and the embedded data has been retrieved completely. The extracting procedure steps are described in Figure 2.

4. Experimental results

The experimental results presented in this section demonstrate the performance of our proposed scheme. To conduct our experiments, we used four 128 × 128 standard grayscale images, ‘‘Baboon,'' ‘‘Lena,'' ‘‘Pepper'' and ‘‘Cameraman.'' These images are shown in Figure 3. A series of pseudo random binary numbers are used as the secret data to be embedded into the cover images.

The well known Peak-Signal-to-Noise Ratio (PSNR) is used as performance measurement criteria, which is classified under the difference image distortion metrics. PSNR is often expressed on a logarithmic scale in decibels (*dB*), it is applied on the stego and the cover images. It is defined as:

$$PSNR = 10 \times log_{10} \frac{255^2}{MSE} (dB). \tag{7}$$

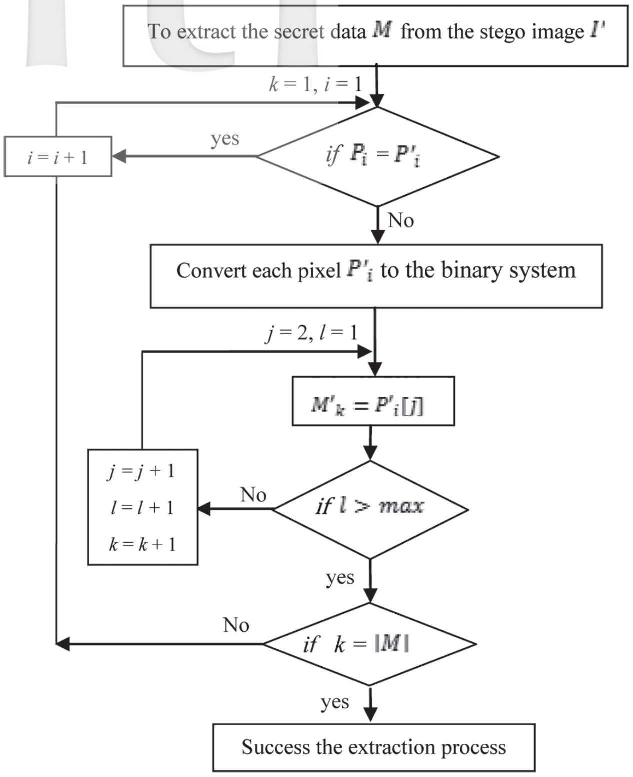
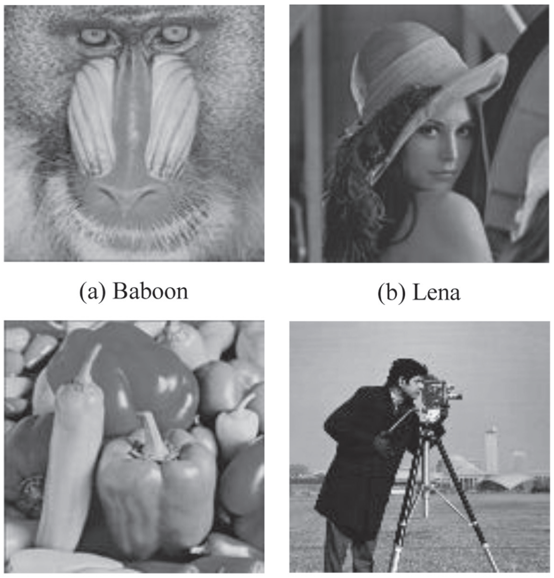


Figure 2 A Block Diagram of Extracting Steps


(a) Baboon

(b) Lena

(c) Pepper

(d) Cameraman

Figure 3 Four 128 × 128 Grayscale Images

Where *MSE* is the mean square error between the cover and stego images. For a cover image whose width and height are *W* and *H*, *MSE* is defined as:

$$MSE = \frac{1}{W \times H} \sum_{i=1}^{w} \sum_{j=1}^{H} (l_{ij} - l'_{ij})^2. \tag{8}$$

Where l_{ij} and l'_{ij} are the pixel values of the cover and stego images, respectively.

Note that, a large PSNR value means that the stego image is most similar to the original image and vice versa. Generally, if the PSNR value is larger than 30 *dB*, then the distortion on the stego image is hard to be detected by human eyes (Chou, Chang, & Li, 2008). Our experimental results show that the proposed method can embed a large amount of secret data while keeping a very high visual quality.

The average of PSNR values are calculated for 6 runs with different randomly selected secret data samples of different lengths. The results of the proposed method for the four grayscale images are shown in Table 1 along with the corresponding results for the classic LSB method, where the *max* value is used as the upper limit criteria for embedding in the proposed scheme, which is also used as the number of LSBs bits for embedding in each pixel in the classic LSB method as shown in Table 1.

We can clearly notice that, the PSNR value of the proposed method is constant although the embedded secret data length is variable, and the embedding criteria is variable as well, because the change in the cover image pixel's value is done only in the first least significant bit. As shown in Table 1, mostly there are significant differences

Table 1 The PSNR Values of the Proposed Method vs. Classic LSB Method

Cover Images	*max* value (no. of LSBs)	PSNR (*dB*)	
		Classic LSB Method	The Proposed Method
Baboon	2	45.11	48.13
	3	39.19	48.13
	4	33.86	48.13
Lena	2	45.13	48.13
	3	38.97	48.13
	4	31.39	48.13
Pepper	2	44.59	48.13
	3	38.84	48.13
	4	33.05	48.13
Cameraman	2	46.62	48.13
	3	37.86	48.13
	4	34.95	48.13

in the PSNR value, makes the proposed technique significantly better compared to the classical LSB.

From Tables 1 and 2, we can conclude that the proposed method performs better if compared to the classic LSB method. To explain, with the same capacity, the proposed scheme provides a higher PSNR, because the embedding targets the identical bits in the cover image pixels with the secret data bits which does not affect the original image quality. Unlike the LSB method, the embedding is done without considering the similarity between the secret data and the cover image bits which definitely will cause changes in the obtained stego image.

Table 2 The Capacity for the Proposed Method and the Classic LSB Method

x^*	**Capacity**	
2	2 bpp	32,768 (bits)
3	3 bpp	49,152 (bits)
4	4 bpp	65,536 (bits)

Note: $^*x$ is the number of LSBs bits, which are inserted into each pixel.

Accordingly, the advantages of the proposed scheme can be summarized as follows: The first advantage appears in the high-quality stego-image visually as well as statistically, where searching for the higher similarity rather than blindly embedding the data minimizes the change in the original image pixel's values. The second advantage is in the distribution of the hidden data, where not all the cover pixels are used for embedding data. Finally, the proposed scheme requires much less computations because its steps are so simple, easy and don't need complex computations such as some techniques which transform image in the frequency domain.

5. Conclusion

Information hiding is a technique for embedding important data into digital media. Steganography, a branch of information hiding, aims to protect important data in transmission. Message capacity and stego image quality are two important criteria in evaluating a steganogarphic method. The basic concept of the proposed method is to use simple LSB substitution. It searches for the similar properties between the secret data bits and the cover image pixels as target for embedding. Our experimental results show that the proposed method provides a stable high stego image quality visually and statistically with high message capacity compared to the classic LSB method.

References

Chan, C.K. & Cheng, L.M. (2004) 'Hiding data in images by simple LSB substitution', *Pattern Recognition*, Vol. 37, No. 3, pp. 469-474.

Chang, C.C., Hsiao, J.Y., & Chan, C.S. (2003) 'Finding optimal least-significant-bit substitution in image hiding by dynamic programming strategy', *Pattern Recognition*, Vol. 36, No. 7, pp. 1583-1595.

Chang, C.C. & Tseng, H.W. (2004) 'A steganographic method for digital images using side match', *Pattern Recognition Letters*, Vol. 25, No. 12, pp. 1431-1437.

Chen, W.J., Chang, C.C., & Le, T.H.N. (2010) 'High payload steganography mechanism using hybrid edge detector', *Expert Systems with Applications*, Vol. 37, No. 4, pp. 3292-3301.

Chou, Y.C., Chang, C.C., & Li, K.M. (2008) 'A large payload data embedding technique for color images', *Fundamenta Informaticae*, Vol. 88, Nos. 1/2, pp. 47-61.

Lin, C.C. & Tsai, W.H. (2004) 'Secret image sharing with steganography and authentication', *Journal of Systems and Software*, Vol. 73, No. 3, pp. 405-414.

Liu, S.H., Chen, T.H., Yao, H.X., & Gao, W. (2004) 'A variable depth LSB data hiding technique in images', *Proceedings of the Third International Conference on Machine Learning and Cybernetics*, Shanghai, China, August 26-29, Vol. 7, pp. 3990-3994.

Lou, D.C. & Liu, J.L. (2002) 'Steganographic method for secure communications', *Computers and Security*, Vol. 21, No. 5, pp. 449-460.

Padmaa, M. & Venkataramani, Y. (2010) 'Zig-Zag PVD -- a nontraditional approach', *International Journal of Computer Applications*, Vol. 5, No. 7, pp. 5-10.

Thien, C.C. & Lin, J.C. (2003) 'A simple and high-hiding capacity method for hiding digit-by-digit data in images based on modulus function', *Pattern Recognition*, Vol. 36, No. 12, pp. 2875-2881.

Wang, C.M., Wu, N.I., Tsai, C.S., & Hwang, M.S. (2008) 'A high quality steganographic method with pixel-value differencing and modulus function', *Journal of Systems and Software*, Vol. 81, No. 1, pp. 150-158.

Wang, R.Z., Lin, C.F., & Lin, J.C. (2001) 'Image hiding by optimal LSB substitution and genetic algorithm', *Pattern Recognition*, Vol. 34, No. 3, pp. 671-683.

Wang, S.J. (2005) 'Steganography of capacity required using modulo operator for embedding secret image', *Applied Mathematics and Computation*, Vol. 164, No. 1, pp. 99-116.

Wu, D.C. & Tsai, W.H. (2003) 'A steganographic method for images by pixel-value differencing', *Pattern Recognition Letters*, Vol. 24, Nos. 9/10, pp. 1613-1626.

Wu, H.C., Wu, N.I., Tsai, C.S., & Hwang, M.S. (2005) 'Image steganographic scheme based on pixel-value differencing and LSB replacement methods', *IEE Proceedings -- Vision, Image and Signal Processing*, Vol. 152, No. 5, pp. 611-615.

Yang, C.H., Weng, C.Y., Wang, S.J., & Sun, H.M. (2008) 'Adaptive data hiding in edge areas of images with spatial LSB domain systems', *IEEE Transactions on Information Forensics and Security*, Vol. 3, No. 3, pp. 488-497.

Zhang, X. & Wang, S. (2004) 'Vulnerability of pixel-value differencing steganography to histogram analysis and modification for enhanced security', *Pattern Recognition Letters*, Vol. 25, No. 3, pp. 331-339.

About the authors

Marghny H. Mohamed received his Ph.D. degree in computer science from the University of Kyushu, Japan, in 2001, his M.Sc. and B.Sc. from Asyut University, Asyut, Egypt, in 1993 and 1988, respectively. He is currently an Associate Professor in the Department of Computer Science, and Vice-President for Student Affairs and Education of the Faculty of Computers and Information Systems, Asyut University, Egypt. His research interests include data mining, text mining, information retrieval, Web mining, machine learning, pattern recognition, neural networks, evolutionary computation, fuzzy systems, and information security. Dr. Marghny is a member of the Egyptian Mathematical Society and Egyptian Syndicate of Scientific Professions. He is a manager of some advanced research projects in Faculty of Computers and Information Systems, University of Asyut, Egypt.

Naziha M. Al-Aidroos received her Bachelor Science degree in computer science in 2003 from Hadhramout University of Science and Technology, Yemen, got here M.Sc. degree in 2009 from Asyut University, Egypt, in computer science. She has worked teacher in Department of Computer Engineering in Faculty of Engineering and Petroleum in Hadhramout University of Science and Technology, Yemen, and she is currently a Ph.D. student in Faculty of Science, Asyut University, Egypt. Her interest subjects are networks, data security, data mining, neural networks, and image processing.

Mohamed A. Bamatraf is an Assistant Professor in Hadhramout University of Science and Technology, Yemen, Faculty of Science, Department of Computer Science. He received his B.S. degrees in Computer Science from Poona University, India, and got his M.Sc. in Computer Science from Osmania University, India, and received his Ph.D. from Asyut

University, Egypt. His Doctoral thesis was about modified data mining techniques and its application in medical diagnosis and intrusion detection. His research areas of interest includes: data and network security, medical informatics, data mining, machine learning, and bioinformatics. His research activities are currently focused on the application of bioinformatics, machine learning, and data security.