Enhanced LSB Image Steganography Method By Using Knight Tour Algorithm, Vigenere Encryption and LZW Compression

Morteza Bashardoost [1], Ghazali Bin Sulong [2] and Parisa Gerami [3]

**[1] Faculty of Computing, Universiti Teknologi Malaysia
Skudai, 81310 Johor, Malaysia**

**[2] Faculty of Computing, Universiti Teknologi Malaysia
Skudai, 81310 Johor, Malaysia**

**[3] Faculty of Computing, Universiti Teknologi Malaysia
Skudai, 81310 Johor, Malaysia**

Abstract

The challenge of steganographic methods is to create a rational balance between the quality of the file and the size of data that can be transferred. In addition, the robustness of the technique and security of the obscure data are the facts that cannot be dissembled. The Least Significant Bit (LSB) insertion approach provides a high degree of visual quality and a large amount of capacity for the concealed data, but the covert message is not well protected in this method. In the proposed method, the secret data is firstly encoded by using the Vigenere encryption method to guarantee the protection of the hidden message. Afterward, the Lempel Ziv Welch (LZW) technique compresses the data to reduce the occupational capacity of the confidential data. Then, by utilizing the extended knight tour algorithm, each bitstream of the data is spread out on the image to increase the robustness of the method. The results show that the proposed method not only improves the security and payload capacity problems of the simple LSB method, but also increases the visual quality of the stego image.

Keywords: *Steganography, LSB embedding technique, Knight Tour algorithm, Vigenere encryption, LZW compression.*

1. Introduction

Data protection and security of the personal information have become a critical issue in the digital world. Therefore, the demand of having a protected method to transfer the confidential data is dramatically increasing. The steganography which literally means "covered writing" [1] is a branch of cryptography and is the art and science of communicating in a way which hides the existence of the communication. In contrast to cryptography which make data unreadable for a third party by implying some encryption methods, steganography emphasize on hiding the existence of message inside another data in such a way that nobody can detect it.

2. Image Steganography Evaluation Parameters

When a large amount of data is embedded into an image the visual specifications of image such as colour and smoothness are altered [2]. Base on this fact that steganography is the process of hiding important information inside a cover data without arising the suspicious, it is very important to specify how the secret is embedded in the image. There are some essential factors that should be considered in image steganography process:

- *Capacity* - The capacity parameter in the steganographic methods refers to the maximum number of bits that can be embedded in a particular cover file with a small probability of revealing by an antagonist.
- *Imperceptibility* - Imperceptibility is defined as the degree of changes in the appearance of the cover data whenever the message is embedded. Since a steganographic system fails if an attacker is able to prove the existence of a secret message, the appearance or format of cover files must remain intact after hiding the secret data.
- *Robustness* - Robustness indicates the distortion amount that the digital cover can tolerate to keep the secret message safe [3]. In a simpler term, the method must guarantee the unity of the message for the receiver even though the stego file is damaged by the performed attacks within the transmission phase.

Security - The security measure denotes the assurance of keeping the secret data unreadable for the adversary when it is extracted by attacks.

3. The Conventional (Simple) LSB Steganography and the Limitations

The Least Significant Bit (LSB) insertion [4-5] is the most common spatial domain technique [6], which consecutively replaces the least significant bit of cover image with the message bits. This method exploits the natural weakness of Human Visual System (HVS) [7] in recognizing the slight difference of colours. The LSB method changes some or all the 8th bit of image's data so that the image's alteration is not perceptible for any human eyes. In like manner, when using a colour image the LSB of each of the red, green and blue components can be used. Therefore, the potential capacity for hiding secret data in a colour image is triple of the same image size in the grayscale mode.

Furthermore, when the data is embedded subsequently to the all bytes of cover image, it would be rather easy to detect and extract the message. A moderately more secure method is to have secret key between the sender and receiver to specify which bytes of image have been used for hiding data [8]. Accordingly, if an adversary receives the image the secret data cannot be extracted without the stego-key.

As it explained above, In spite of having the highest payload and also high degree of quality and imperceptibility, LSB algorithm is not highly secure against statistical attacks and the protection of hidden data is not guaranteed. To say in other terms, by extracting the data from cover image, it would be rather easy to find the original message [4]. Taking all above into consideration, is there any way to surge the security of LSB method besides having significant quality and imperceptibility?

4. Previous Attempts

Many algorithms that work in the spatial domain utilize the LSB technique or any of its derivatives as the algorithm for information hiding. But these methods cannot resist against some type of statistical analysis such as RS [9] or Sample Pairs [10], even if partially disguised in the amount of information hidden. The problem originates from the fact that embedding the secret data in the cover image led to a distortion that is not perceptible for the human eyes, but is detectable by statistical analysis.

The Optimal LSB insertion is an enhanced LSB method, which performs an adjustment process to find the optimal pixels to improve the stego-image quality. Indeed, three candidates are selected for the pixel's value to compare which one has the closest value to the source pixel value when the secret data embedded in. Then the best candidate is called the optimal pixel and used for hiding the secret data [4].

The Pixel Value Differencing method (PVD) [11] utilizes the characteristic of human visual system to extend the capacity of the image for hiding data beside the high level of imperceptibility. Since smooth areas and edge areas have different payload capacity, the edge area can embed more secret data inside [12]. In fact the distortion tolerance degree of a smooth area is generally less than an edge area. Furthermore, in PVD technique the characteristics of image blocks remain constant because this method does not change any smooth area to the edge area and vice versa.

Applying randomization concept to LSB method is one more LSB improved method, which works on the basis of the theory that the reaction of human eyes to Red, Blue and Green is different [8]. According to the brightness formula (I = 0.3R + 0.59G + 0.11B), human eyes are most sensitive to the green, the next is to the red, and the least is to the blue. Therefore, the different least bits of brightness components can be utilized for hiding data. Furthermore, by taking advantage of the less eyesight relevance to the lower bits, the data to be hidden are embedded into the lower (first few least) bits of each pixel.

5. Proposed Steganography Method

The general architecture of the proposed steganography method is designed and implemented. As it is depicted in the Figure 1, the whole process is composed of two main phases, which are Embedding phase and Receiving phase.

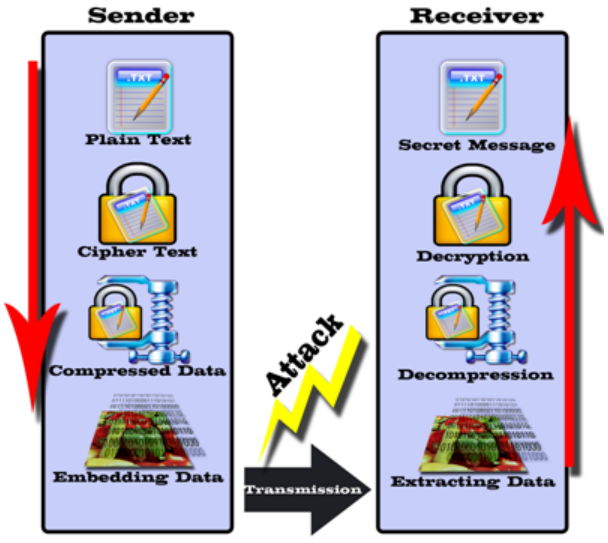


Fig. 1 Proposed Framework.

5.1 Embedding Process

This phase includes all the activities that must be carried out to hide and protect the secret data inside the cover image. The sender uses some algorithms to encode and compress the data and then embeds the bit stream into the image. Moreover, the secret key is defined as the first position of the bit stream within the image. This key is identified just for the sender and receiver. The sending process consists of following procedures:

- *Encryption* - In the first step of the embedding phase, the plain text will be encrypted using the Vigenere table [13]. There are several encryption methods that can be applied to encrypt the data, but in this situation, we need a method that does not produce a cipher text longer than the plain text. Furthermore, among the desired encryption methods, Vigenere table, which is a symmetric encryption [14] technique and maps each input character into exactly one character for output, is more secure than similar methods. The biggest advantage of Vigenere table over the other symmetric encryption methods is that based on the specified secret key, it produces different outputs for a certain input character.
- *Compression* - Compressing the message can be a suitable solution for the limited payload space of the host image. Shortening the message size not only increases the capacity of the cover image, but also reduces the probability of discovering the message inside the host image. The LZW compression method [15] is employed effectively to diminish the size of the message. LZW creates a table to replace the repetitive succeeding characters with a binary code. This table, which is known as dictionary, will be sent to the recipient the end of the compression process to be used for extracting original secret message.
- *Embedding* - The embedding algorithm is the most prominent part of the steganographic methods. In fact, it defines which pixels of the image should be changed and also in what order they will be altered with the secret data.

The "Knight tour" algorithm is a suitable technique to formulate the sequence of the secret bit stream within the image pixels. The advantage of the knight tour method over the PRNG technique [16] is that, it is a self-developed algorithm based on the knight tour mathematical problem [17] and it is almost unknown for the unintended receivers. By considering the image as an extended chessboard, we can have an algorithm, which determines the path of the knight within the image.

The solution of the "Knight Tour" problem divides the chess board into the blocks with the size of 4x4 squares. Also, it considers four groups of four squares in each block namely "Right Diamond", "Left Diamond", "Right Square" and "Left Square". The main rule of the surfing is to complete all the squares within the chessboard on each group and then move to the next group of squares. The Figure 2 displays a block and its constitutive groups of squares:

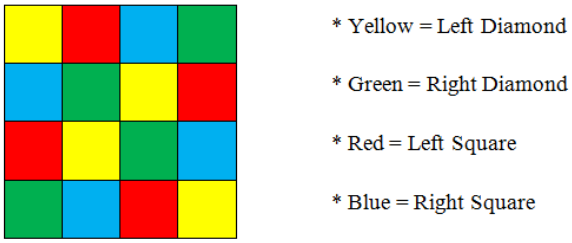


Fig. 2 Block of Pixels, Containing Four Groups of Squares.

The embedding algorithm of the proposed steganographic method is based on the described "Knight Tour" algorithm. However, it has the larger board and also it is not necessary to pass all the squares continuously. If the size of the image is divisible by four, the algorithm will cover all the pixels and otherwise, the extra columns or rows (which are less than 4) will be unusable. The proposed embedding technique follows the following steps:

a) Consider the image's width and height divisible by four (Ignore the extra pixels).
b) Divide the image into 4x4-pixel blocks.
c) Go to the first pixel, which has been specified by the stego-key and start with one group (colour) and traverse all the blocks.
d) To move from one block to the next one, all the 4 squares must be traversed.
e) If the movement for one group (colour) has finished, start with the next group.
f) Repeat the steps to traverse all the pixels.

Replacement – When the sequence of the target pixels is defined in the previous step, now it's the time to replace the least significant bits of the image pixels with the bit stream of the secret message.

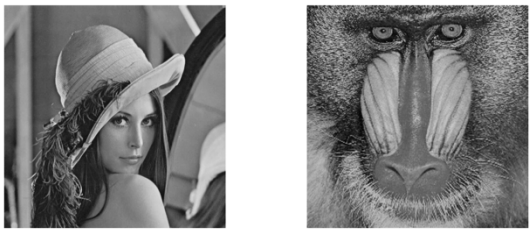
5.2 Extraction Phase

On the other side of the communication line, the receiver should be able to comprehend the secret data within the Stego-image. Therefore, another procedure is required to recover the content of the message and restructure it.

First of all, base on the stego-key and the extracting algorithm (the same as sender side) the bits of the secret message are obtained to compose a compressed data. Then the unzipping algorithm will generate the encrypted data and finally, by using Vigenere table the plain message will be revealed.

6. Experimental Result

The proposed method is implemented by using MATLAB tools and images of Lena and Baboon, which are illustrated in Figure 3, being used in 8-bit grayscale mode for testing. In addition, these images are applied in both sizes of 256 × 256 and 512 × 512. Furthermore, the secret messages, which are used in the implementation phase, are arbitrary text with different size of 512, 1024 and 2048 Bytes.



(a) (b)

Fig. 3 Standard Photos of Lena (a) and Baboon (b).

As mentioned earlier, the steganographic methods are evaluated based on some key factors, which are imperceptibility, capacity, robustness and security. In this section, the proposed method is examined with respect to each essential metric, and then is analysed precisely to explain the advantages of the proposed method.

6.1. Imperceptibility

This criterion determines the quality of the stego image and specifies how much the output is similar to the cover image. A greater value for the PSNR quality metric indicates a lower degree of distortion for the generated image.

Table 1 presents the results of the both Simple LSB method and proposed technique to embed different size of payloads in sample pictures of Lena and Baboon by the size of 256 × 256. The outcome values indicate that the enhanced LSB method produces higher PSNR values in all tested cases. Furthermore, the PSNR value differs for the different images that are tested with each method. However, by increasing the amount of payload in each image a significant fall appears in the PSNR value.

Table 1: Payload Size - Simple LSB Method and Enhanced LSB Method.

	Simple LSB				Enhanced LSB			
	Lena		Baboon		Lena		Baboon	
Message size (Bits)	32768	65536	32768	65536	32768	65536	32768	65536
Embedding rate	0.0625	0.125	0.0625	0.125	0.0625	0.125	0.0625	0.125
PSNR (dB)	57.4995	54.5289	57.6139	54.5678	59.8601	56.8745	59.9034	56.9154

6.2 Capacity

If all the pixels of the 8-bit grayscale image are used for embedding the secret data, the maximum size of secret data, which can be embedded in the image by using conventional LSB method, is calculated from the Formula 1:

$$\text{Capacity} = (\text{Image width} * \text{Image height})/8 \quad \text{(Bytes)} \quad (1)$$

It is expected that the maximum size of payload increases due to LZW compression. Based on the LZW algorithm, as much the size of data growth the compression ratio also goes up. Moreover, the maximum size of payload depends on the size of the message and the combination of the characters. Therefore, it cannot be calculated easily and based on the compression trend it can be estimated. The growth in the rate of compression can be clearly observed in the Figure 4:

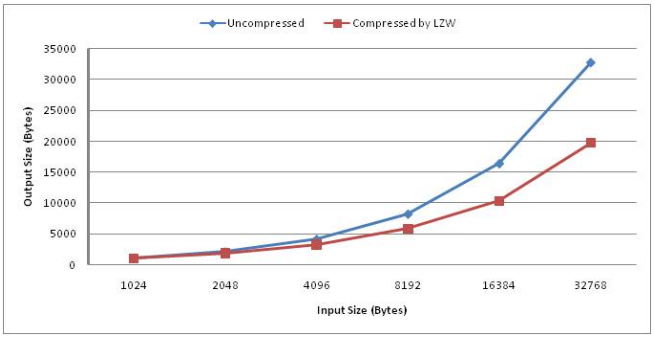


Fig. 4 Compression Ratio in Enhanced LSB Method.

The Table 3 indicates that the maximum size of a message, which can be embedded into the LSB bit of an 8-bit grayscale image, can be increased when the data is compressed. In this calculation, we consider all the pixels of the image are used for embedding confidential data:

Table 2. Maximum Payload Size – Simple and Enhanced LSB methods.

	Simple LSB			Enhanced LSB		
Image dimension (pixels)	256 × 256	512 × 512	1024 × 1024	256 × 256	512 × 512	1024 × 1024
Maximum size of secret data (Bytes)	8192	32768	131072	11625	54540	234058

6.3 Security

The Chi-square statistical attack is applied on the suspected images, to check whether they are conveying any hidden data or not. This attack is based on the distribution probability of zeros and ones over the image. The existence frequency of each of the two pixel values in each POV differs from the mean of the POV.
Figure 5 and Figure 6 illustrate the result of Chi-square attack on the stego image, which carries 4 and 8 kilobytes of data by using Simple LSB method. As it seems in the charts, the existence of the message can be easily detected when the probability trend falls dramatically from around one (hundred percents) to zero:

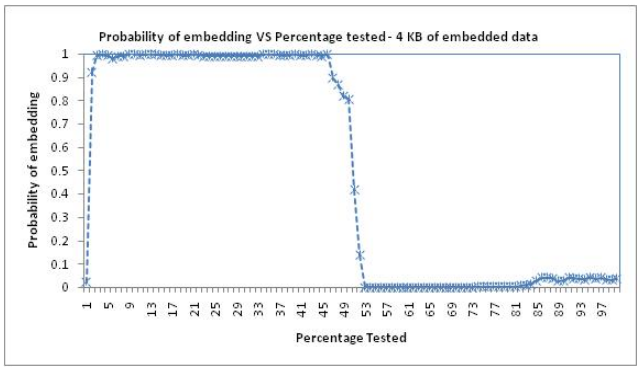


Fig. 5 Chi-square result of Simple LSB method for embedding 4KB data.

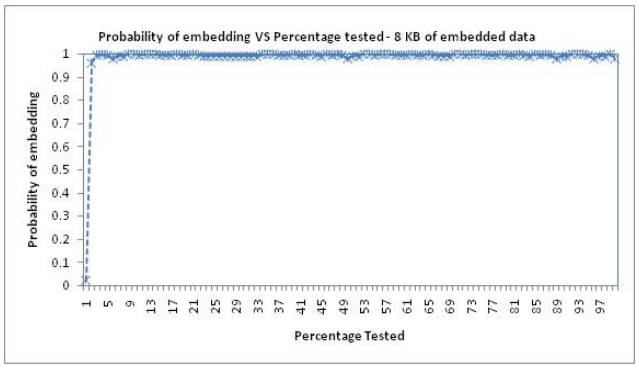


Fig. 6 Chi-square result of Simple LSB method for embedding 8KB data.

When the proposed approach is used for embedding data into the same cover image the results are noticeably different. The Chi-square diagram almost detects no embedded data in the tested images; however, the payload size in the Figure 7 is 4 KB and in the Figure 8 is 8 KB, and it is expected to see 50 and 100 percents of probability in these charts. The deviation from zero, just are visible in few points of probability trend and especially, these values are not even more than one percent. Therefore, the results demonstrate the stability of the Enhanced LSB method against Chi-square statistical attack:

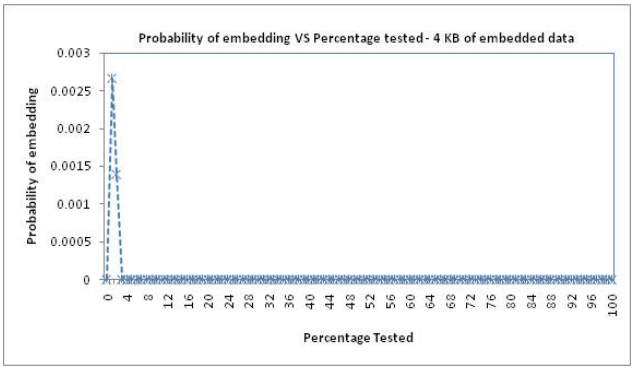


Fig. 7 Chi-square result - Enhanced LSB method (embedding 4KB data).

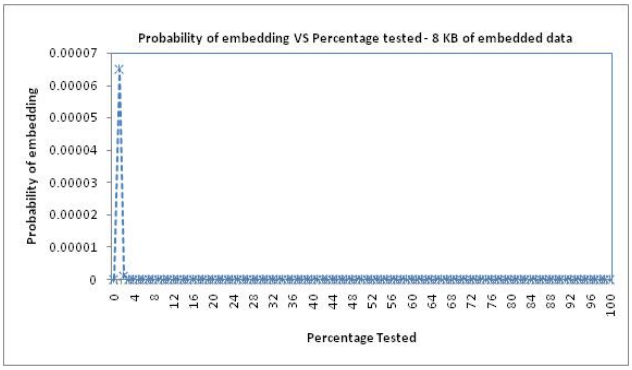


Fig. 8 Chi-square result - Enhanced LSB method (embedding 8KB data).

To have a more precise evaluation of our method in terms of security, the LSB layer visual attack applied on the produced stego-image. Since the LSB-based steganographic methods produce high quality stego images, the artifacts of the image are not observable for human eyes. In the Enhanced LSB attack image is converted to its eight composing layers and the layer number zero will be probed which contains the least significant bits of each pixel.
As it seems in the Figure 9, the LSB layer of the cover image is composed of some arbitrary black and white pixels without any recognizable pattern. However, when the secret message is embedded using Simple LSB

technique, the Layer zero shows some artifacts inside. The vertical strip pattern has appeared when 50 percent of the pixels (Figure 10 a) and all pixels (Figure 10 b) are modified by hidden message. Furthermore, when the proposed algorithm is exploited, the LSB layer seems entirely innocent. The result of the enhanced LSB method is depicted in Figure 11:

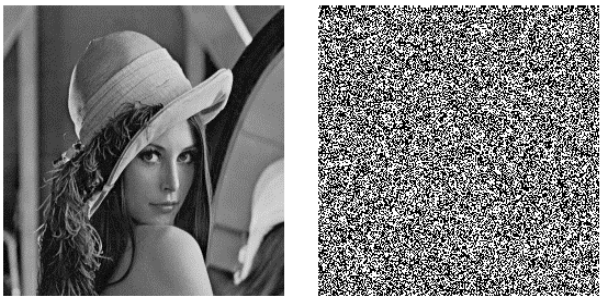


Fig. 9 Lena Test Image (a) and its Least Significant Bit Layer.

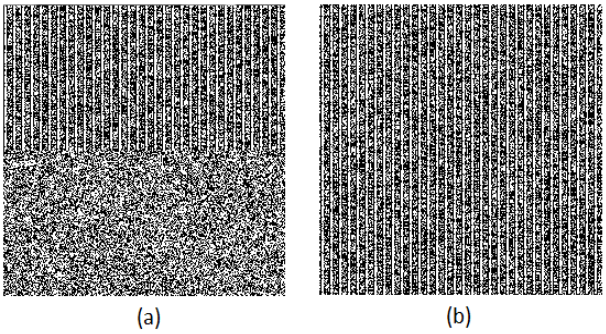


Fig. 10 The Result of Simple LSB method: Layer Zero with 50% of Hidden Data (a) and 100% of Hidden Data (b).

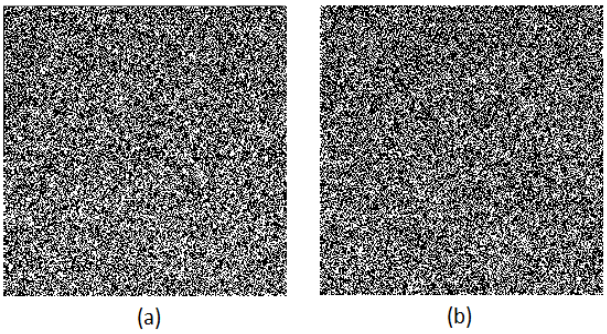


Fig. 11 The Result of Enhanced LSB method: Layer Zero with 50% of Hidden Data (a) and 100% of Hidden Data (b).

7. Conclusion

In this work, we decided to fix the weakness of Simple LSB system by providing some enhancements. The Enhanced LSB method utilizes three fundamental improvements specifically Knight Tour embedding algorithm, Vigenere encryption and LZW compression.
The process starts with the encoding the confidential information by using Vigenere encryption technique. Both of the sender and receiver have a secret key which is used in encryption and decryption phases. Afterward, the LZW compression technique reduces the size of encrypted data to improve the payload capacity. Clearly, as much the length of input data increases, the rate of compression surges, as well. Finally, the generated bitstream are embedded into the image in the positions which are defined by the proposed embedding algorithm. The aforesaid embedding method is extended form of Knight Tour algorithm and provides the maximum number of pixels to hide the secret message.
As it was expected theoretically, the satisfactory results were achieved when the method was implemented. Results proved that the Enhanced LSB method saves up to forty percent of capacity because of exploiting compression technique. Therefore, little number of pixels of the image will be probably modified and consequently the quality of the stego image will be improved. In addition, smaller amount of data will be distorted whereas the third party applies active visual attacks on the Stego image. Finally, the possibility of extracting the content of hidden data reduces significantly, when the private message becomes encrypted.

8. Future Works

In this study, we tried out to increase the security level of Simple LSB method and because of using the compression phase in the proposed method, not only the capacity of payloads increased, but also the quality and robustness are enhanced. However, the proposed method can be improved if the following criteria are considered in the future works:

- Examine the stability of the method against other statistical attacks.
- Improve the robustness of the method by taking other bits of the host image into the consideration.
- Enhance the method in terms of the integrity. That means whenever the secret data being modified within the transmission channel the receiver must realize that is a fake message.

References

[1] A. Cheddad, *et al.*, "Digital image steganography: Survey and analysis of current methods," *Signal Processing,* vol. 90, pp. 727-752, 2010.

[2] R. J. Anderson and F. A. P. Petitcolas, "On the limits of steganography," *Selected Areas in Communications, IEEE Journal on,* vol. 16, pp. 474-481, 1998.

[3] H. Tariq Al, *et al.*, "A testbed for evaluating security and robustness of steganography techniques," in *Circuits and Systems, 2003 IEEE 46th Midwest Symposium on*, 2003, pp. 1583-1586 Vol. 3.

[4] C. K. Chan and L. M. Cheng, "Hiding data in images by simple LSB substitution," *Pattern Recognition,* vol. 37, pp. 469-474, 2004.

[5] S. Dey, *et al.*, "An LSB Data Hiding Technique Using Prime Numbers," in *Information Assurance and Security, 2007. IAS 2007. Third International Symposium on*, 2007, pp. 101-108.

[6] A. Daneshkhah, *et al.*, "A More Secure Steganography Method in Spatial Domain," in *Intelligent Systems, Modelling and Simulation (ISMS), 2011 Second International Conference on*, 2011, pp. 189-194.

[7] C. Yi-zhen, *et al.*, "An adaptive steganography algorithm based on block sensitivity vectors using HVS features," in *Image and Signal Processing (CISP), 2010 3rd International Congress on*, 2010, pp. 1151-1155.

[8] J. V. Anand and G. D. Dharaneetharan, "New approach in steganography by integrating different LSB algorithms and applying randomization concept to enhance security," presented at the Proceedings of the 2011 International Conference on Communication, Computing, Rourkela, Odisha, India 474-476, 2011.

[9] L. Xiangyang, *et al.*, "Detecting LSB steganography based on dynamic masks," in *Intelligent Systems Design and Applications, 2005. ISDA '05. Proceedings. 5th International Conference on*, 2005, pp. 251-255.

[10] S. Dumitrescu, *et al.*, "Detection of LSB steganography via sample pair analysis," *Signal Processing, IEEE Transactions on,* vol. 51, pp. 355-372, 2003.

[11] H. C. Wu, *et al.*, "Image steganographic scheme based on pixel-value differencing and LSB replacement methods," *Vision, Image and Signal Processing, IEE Proceedings -,* vol. 152, pp. 611-615, 2005.

[12] J. Liping, *et al.*, "A Further Study on a PVD-Based Steganography," in *Multimedia Information Networking and Security (MINES), 2010 International Conference on*, 2010, pp. 686-690.

[13] V. T. Dennie, "Cryptographic techniques for computers: Substitution methods," *Information Storage and Retrieval,* vol. 6, pp. 241-249, 2007.

[14] S. Aruljothi and M. Venkatesulu, "Symmetric Key Cryptosystem Based on Randomized Block Cipher," in *Future Information Technology (FutureTech), 2010 5th International Conference on*, 2010, pp. 1-5.

[15] J. Kärkkäinen, *et al.*, "Approximate string matching on Ziv–Lempel compressed text," *Journal of Discrete Algorithms,* vol. 1, pp. 313-338, 2003.

[16] I. M. Sobol and Y. L. Levitan, "A pseudo-random number generator for personal computers," *Computers & Mathematics with Applications,* vol. 37, pp. 33-40, 1999.

[17] I. Parberry, "An efficient algorithm for the Knight's tour problem," *Discrete Applied Mathematics,* vol. 73, pp. 251-260, 1997.

Morteza Bashardoost received his bachelor's degree of software engineering from Islamic Azad University, Lahijan, Iran in 2005. Afterward, he obtained his master of computer science from Universiti Teknologi Malaysia (UTM), Johor, Malaysia in 2012. He is currently a Phd student of computer science in Universiti Teknologi Malaysia. His areas of research interest include digital image processing, pattern recognition, face recognition, image steganography and watermarking.

Ghazali Bin Sulong received his Ph.D. and M.Sc. degrees in computing from University of Wales, Cardiff, United Kingdom in 1989 and 1982 respectively, and B.Sc. degree in statistic from National University of Malaysia (UKM), in 1979. Currently, he is a professor and principle researcher in image processing in Universiti Teknologi Malaysia (UTM).

Parisa Gerami studied computer science in Zarandieh Institute of Higher education (B.Sc) in 2009 and received her master degree from university technology Malaysia in computer science (information security) in 2012. Her research interests are Image Processing, Computer Vision, Cryptography, Steganography, Watermarking, Security, Networking Security, Wireless Sensor Network and Wireless Multimedia Sensor Network.