Using Image Steganography for Providing Enhanced Medical Data security

Muhammad Arslan Usman[1], *Student Member IEEE* and Muhammad Rehan Usman[2], *Student Member IEEE*
Department of IT Convergence Engineering
Kumoh National Institute of Technology (KIT)
Gumi, Republic of Korea (South Korea)
arslanusman@msn.com[1], rehanusman@gmail.com[2]

***Abstract*—Recently, image steganography is being considered as an alternative method for securing medical data to avoid medical related cybercrimes. This paper proposes a new image steganography approach for securing medical data. Swapped Huffman tree coding is used to apply lossless compression and manifold encryption to the payload before embedding into the cover image. Additionally, only edge regions of the cover image are used to embed the secret data which offers high imperceptibility. The results show that the proposed method ensures confidentiality and secrecy of patient information while maintaining imperceptibility.**

Keywords—Edge detection, Huffman coding, image steganography, least significant bit, medical data security.

I. INTRODUCTION

With recent and rapid advancements in communication technologies, digital signals can be transmitted over the internet with convenience [1]. These advancements have brought many advantages but at the same time there are several hazards and risks that need to be considered as well. Technologies such as telemedicine are emerging day by day [2] and ensuring medical data security is becoming a challenge. Recently, capturing medical data has appeared as a major cybercrime. If such sensitive data is stolen or captured, then it can result in violation of basic patient rights. Confidentiality in medical reports must be kept intact in order to ensure trust among patients and health care institutions. Electronic health records (EHR) are stored in large databases of medical institutions, in which patient's health records are kept [3]. These records may include every sort of sensitive information starting from patient's personal data, vital signs, diagnosis reports to laboratory reports etc. This data serves as a medical history for the convenience of doctors and patients. These medical records are shared through modern communication systems which include variety of networks such as local or wide area networks. Out of all the medical information kept in EHRs, 90% data is comprised of medical images. Medical images, such as X-ray, endoscopy images and videos, MR (magnetic resonance) images etc., are stored, handled and transmitted using the Digital Imaging and Communications in Medicine (DICOM) standard. The patient information in DICOM file must be kept confidential to avoid any kind of tampering of patient's data, illegal copying and to guarantee copyright protection [3]. This confidentiality requires medical data to be secured in every way possible. Medical data security provides protection to patients' rights and at the same time ensures the responsibility of health professionals. Firewalls, virtual private networks (VPN), encryption and cryptography techniques help in ensuring medical data security. Unfortunately, hackers can easily bypass firewalls and moreover encryption techniques do not fully guarantee medical data security. Hiding secret or confidential information in different mediums, such as images, audio signals etc., is gaining popularity among researchers. Apart from contemporary methods, steganography is one way to hide secret information and can be considered as an alternate way to ensure medical data security. Image steganography is a method that embeds an image, called as secret image, into another image called as a cover image. Image steganography in general can be classified in two domains, namely the transform and spatial domain. There are three requirements for a good image steganography approach as shown in Fig. 1. But there exists a tradeoff between these three requirements. If the capacity is increased, then imperceptibility is somehow compromised and vice versa. The transform domain methods in image steganography is immune to attacks like compression, filtering etc., but they have limited capacity, offer high complexity and computational time. On the other hand, spatial domain offers high capacity and robustness and, they have less computational time and complexity. The most common method in spatial domain is the least significant bit (LSB) replacement method. In this method, the LSBs of a pixel in a cover image are replaced with MSBs of secret image. Examples for image steganography methods in spatial domain can be found in [4, 5].

In this paper, a novel image steganography method is proposed which not only fulfills the three requirements of a good steganography approach, but offers manifold data encryption for medical images. The study aims to; (i) offer imperceptibility by using only the edge regions of cover image for data embedding, (ii) ensure high capacity by applying lossless compression to the secret images, and (iii) encrypt and secure the compressed data by applying swapped Huffman tree encoding (SHT).

Rest of the paper is organized as follows; a brief survey of related works is presented in Section II, Section III contains brief description of the edge detection method used in the proposed work followed by details of the proposed method with systematic explanation of the data embedding and extraction algorithm. Section IV covers the experimental evaluation of the proposed method followed by conclusion in sections V.

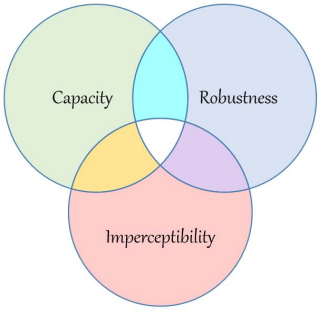

Fig. 1. Requirements of a good image steganography approach

978-1-5386-4790-5/18/$31.00 ©2018 IEEE

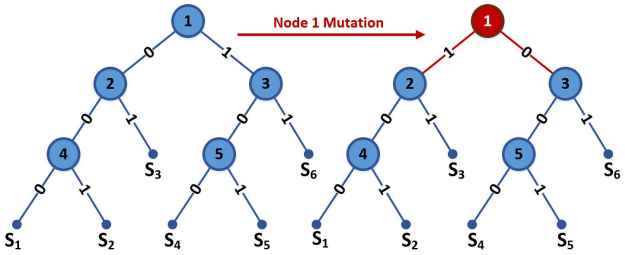

Fig. 2. The concept of mutation

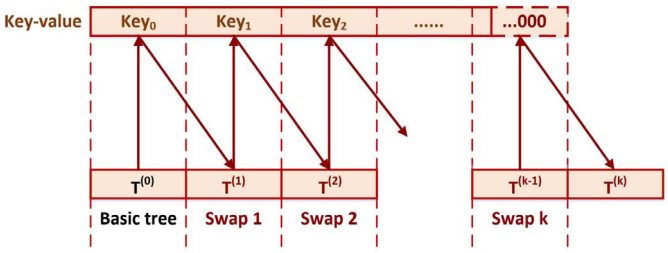

Fig. 3. The swapping concept in SHT.

II. Related Work

The authors in [7] have presented a method for visually imperceptible data hiding in Positron emission tomography (PET) images by considering ECG images as secret images. They have used multi-resolution wavelet decomposition in selected regions of PET images which are considered to be imperceptible to human visual system (HVS). In [8], a blind reversible watermarking approach for medical images based on histogram shifting in wavelet domain was proposed. The authors have used integer wavelet transform to map the integer host image components to integer wavelet coefficients. Another medical data hiding approach using lossless compression was proposed in [9]. The authors performed the embedding when the original image is compressed using the embedded zero-tree wavelet (EZW) algorithm. The authors in [10] presented a watermarking scheme that combines data compression, encryption and watermarking techniques and image moment theory applied to radiological medical images. In their proposed framework, they have used *DICOM* data as a watermark to embed in medical images. A reversible data hiding scheme for medical images was proposed in [11] which is focused on dividing the image into tiles and shifting the histograms those tiles between their minimum and maximum frequency. Secret data is then inserted at the pixel level with the largest frequency to maximize the payload hiding capacity. A review of several state-of-the-art watermarking techniques for medical images can be found in [6] [12, 13].

III. Proposed Work

This section contains the details of our proposed method while briefly describing the edge detection in images and explains the Canny algorithm that we have used in our method. Further, the section includes description about swapped Huffman tree coding (SHT) and encryption starting with the basic Huffman coding. Also, the data embedding, and extraction process of the proposed image steganography method is explained in detail. The security analysis of SHT coding is not the scope of this paper and is provided in its corresponding publication which is cited in this paper.

A. Edge Detection

It is observed that the pixels in the edge regions present more random characteristics than the smooth regions in an image. Also, it is known that HVS is more sensitive towards slight changes in the smooth regions in an image as compared to the edge regions [14]. Visual impairments and noticeable artifacts are hard to find in the edge regions if secret data is embedded in them. In this paper, we have used edge regions of medical images to embed the secret data. The Canny algorithm is widely used for edge detection especially in medical images [15]. It has two parameters that can be adjusted per the requirements. Firstly, it smoothens the image through a Gaussian filter to reduce the noise. So, the first adjustable parameter decides the size of the mask of Gaussian filter. The second parameter includes two thresholds, a high and a low one, which can be adjusted flexibly.

The next subsection contains details about the SHT algorithm followed by the data embedding and extraction algorithm of our proposed method.

B. Swapped Huffman Tree Coding

Huffman coding, which is a type of lossless compression, is a type of optimal prefix code. A detailed description of Huffman coding can be found in [16]. Application of basic Huffman coding is mainly compressing the data without the loss of information. Encryption in the compressed data was introduced by researchers in later years. Such techniques include mutated Huffman table (MHT) [17] and chaotic Huffman table (CHT) [18]. Authors in [19] showed that MHT is vulnerable to attacks, so it lost its significance. SHT appeared later which is more reliable than its counterparts [20].

Encryption and compression are performed together in SHT where the lossless compression is based on Huffman coding [20, 21]. Encryption is performed on the coded data (a binary bit stream) and the technique used is called mutation. As shown in Fig. 2, mutation is done by toggling between the branch labels of the tree node, i.e., '0' to '1' and vice versa. SHT is considered more reliable than MHT technique because it also uses encryption key swapping along with the mutation for encryption. The encryption key swapping mechanism includes multiple mutations by dividing the total key stream into further key segments according to the number of nodes in a tree as shown in Fig. 3, where T^k represents the number of tree mutation. The total number of bits in each segment are decided as per the number of nodes in the Huffman tree. Each key segment is applied to a tree in a way that if '1' occurs then the

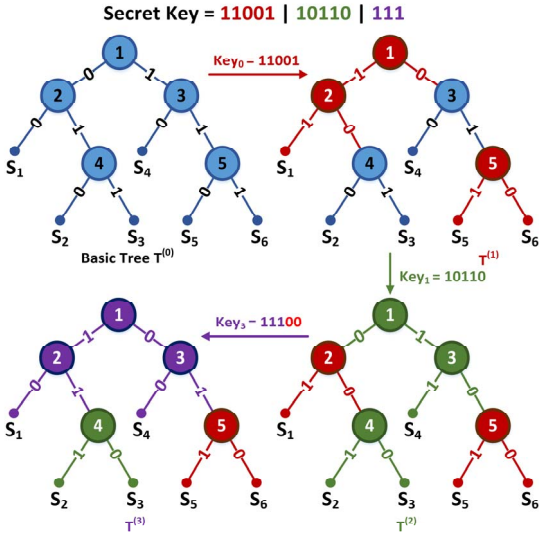

Fig. 4. Example: Swapped Huffman Tree (SHT) Encoding.

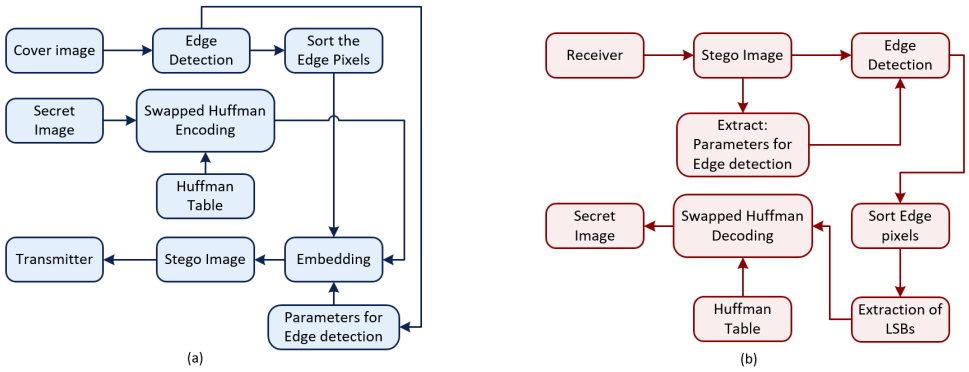

Fig. 5. a) Flowchart for embedding the secret data. b) Flowchart for extracting the secret data.

bit is toggled or mutated otherwise it is kept as it is. As the length of the secret key has no restrictions so it is resilient to brute force attacks. Furthermore, it does not impact the compression ratio after the encryption. Using Fig. 4, following example is provided to give a better understanding of the SHT process.

1. Define the Huffman tree. Let's suppose the tree $T^{(0)}$ in Fig. 4.
2. Count the number of nodes M in the Huffman tree, e.g., in Fig. 4, $M=5$.
3. Define the mutation *key*, let's suppose the *key* has x number of bits, then the length of the *key* can be $M < x < \infty$, e.g., in Fig. 4, the length of the chosen *key* is 13 bits, i.e., 1100110110111.
4. Divide the key in step 3 in to further key segments in a way that the number of bits in each segment must be equal to M. The key segments are denoted by key_n where $n = \{0, 1, \ldots, k\}$ is the total number of segments. If the length of the k^{th} segment is less than M then add zeros to the right of the key segment, e.g., in Fig. 4, the value of M is 5 so the division of the *key* into key segments is 11001|10110|111. Here we see that the segment key_3 has three bits which is less than M so adding two 0s to the right will yield key_3 = 111 00.
5. After the definition, apply the first segment key_0 to the Huffman tree in step one $T^{(0)}$ to generate the mutated Huffman tree $T^{(1)}$. Then apply key_1 to $T^{(1)}$ for further mutation to generate $T^{(2)}$ and so on until the last segment of the defined *key* in step 3 is reached, e.g., in Fig 4, the key segments; key_0 = 11001 is applied on $T^{(0)}$ to generate $T^{(1)}$, key_2 = 10110 is applied on $T^{(1)}$ to generate $T^{(2)}$ and key_3 = 11100 is applied on $T^{(2)}$ to generate $T^{(3)}$.
6. The final tree $T^{(k)}$ with the k^{th} number of key segment will be the final mutated tree and will be highly secure because of multiple layers of mutation. This manifold encryption is only possible because of the key segmentation. This final tree will be used to generate the Huffman table for transmission, e.g., in Fig. 4, the k^{th} number of key segment is key_3=11100 applied to $T^{(2)}$ to generate the final mutated tree $T^{(3)}$.

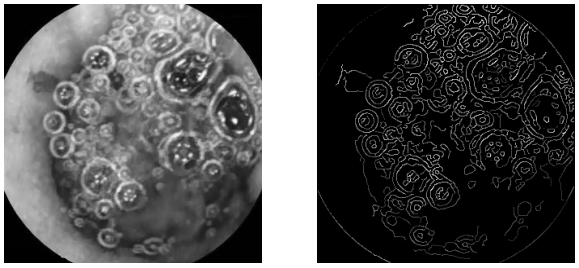

Fig. 6. Results of Canny edge detection.

The final step is to generate the Huffman table from $T^{(k)}$ after the SHT process from step 1 to step 6 is completed.

In this way, images are compressed and encrypted at the same time using the same technique without a loss in the information bits. The secret 2D image with $L=W\times H$ dimensions is compressed and encrypted into a 1D binary bit stream using SHT by assigning a binary code word to each intensity value of the secret image. The total length of the binary bit stream is less than L.

The transmitter and receiver side, both should have the swapped and mutated Huffman table, from step 7, to encode and decode the images respectively. So, the swapped and mutated Huffman table along with the information, which represents the number of bits in each segment of the secret encryption key, must be embedded in the cover image in addition to the compressed and encrypted image data. In this paper the receiver and the transmitter are assumed to be perfectly synchronized, i.e. the receiver has the information about the secret encryption key that is used to encrypt the images.

The flowcharts for embedding and extraction of the secret image are given in Fig. 5.

IV. Experimental Evaluation

In this paper, we have used two different types of medical images for experimental evaluation of the proposed work. Wireless capsule endoscopy (WCE) images, with 320×320 dimensions, are used and MRI scans of brain, with 512×512 dimensions, are used. For simplicity, we have used 8-bit gray scale cover images. A 128×128, 8-bit grayscale image is used as the secret image.

Fig. 6 shows the detected edges through Canny edge detection method. The size of the Gaussian filter was set to be 1.5, and the low and high thresholds were kept to be 0.05 and 0.25 respectively. We have used Canny edge detection in our work because it gives the highest quality in stego images in terms of PSNR. The capacity of our proposed method is dependent on the number of edge pixels that are detected. The detected edge pixels in Fig. 6 are used to insert the secret image (which is the medical data in our case) after it is encrypted and compressed using the SHT algorithm.

Fig. 7 shows the histogram of different cover and stego images. It can be observed that the retrieved image does not contain noticeable impairments that can be detected by naked human eye. There are slight variations in the peaks of the histograms of stego images as compared to the cover images, but this can happen due to addition of random impulse noise during the transmission and reception process. So, observing the

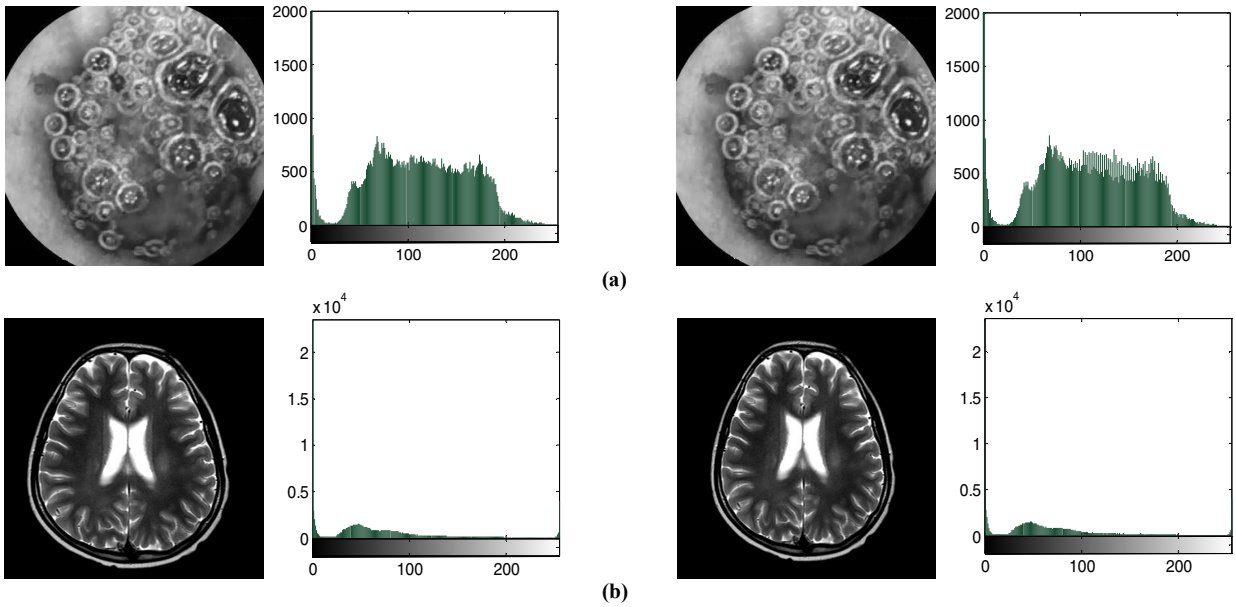

Fig. 7. The images in first, second, third and fourth column represent cover images, histograms of cover images, stego images and histograms of stego images respectively. **a)** From left to right → A WCE image containing bleeding due to Angiodysplasia, its corresponding histogram, stego image and its corresponding histogram. **b)** From left to right → A brain MR image, its corresponding histogram, stego image and its corresponding histogram.

histograms, it can be inferred that the visual impairments in the stego images, due to the addition of secret data, are not easily detectable through naked human eye.

V. CONCLUSION

Medical data can easily be stolen or captured during storing, transmission or reception through a network and internet. Such cybercrimes can be avoided if medical data security is ensured. The proposed image steganography approach shows that the medical data or patient's data can be secured and encrypted while maintaining the quality and imperceptibility in stego images. Swapped Huffman tree encoding is applied to provide manifold encryption to the medical data. Comparing the histograms of the stego image and the cover image, it can be observed that imperceptibility of the hidden data is maintained.

REFERENCES

[1] M. A. Usman, S. Y. Shin, M. Shahid, & B. Lövström, "A no reference video quality metric based on jerkiness estimation focusing on multiple frame freezing in video streaming," IETE Technical Review, Vol. 34, No. 3, pp: 309 – 320, May 2017.

[2] M. A. Usman, M. R. Usman, & S. Y. Shin, "Quality assessment for wireless capsule endoscopy videos compressed via HEVC: From diagnostic quality to visual perception." Computers in Biology and Medicine Vol. 91, pp: 112-134, 2017.

[3] L. Q. Kuang, Y. Zhang, & X. Han, "Watermarking image authentication in hospital information system," *Information Engineering and Computer Science*, ICIECS, pp. 1–4, 2009.

[4] C. K. Chan, & L. M. Chen, "Hiding data in images by simple LSB substitution," *Pattern Recognition*, vol. 37, no. 3, pp. 469–474, 2004.

[5] E. Kawaguchi, & R. O. Eason, "Principle and applications of BPCS-Steganography," *In Proceedings of SPIE- The International Society for Optical Engineering,* vol. 3528, pp. 464-473, 1999.

[6] A. Cheddad, J. Condell, K. Curran, & P. Mc Kevitt, "Digital image steganography: Survey and analysis of current methods," in *Signal processing,* vol. 90, no. 3, pp: 727-752, 2010.

[7] M.S. Nambakhsh, A. Ahmadian, & H. Zaidi, "A contextual based double watermarking of PET images by patient ID and ECG signal," in *Comput. Methods Progr. Biomed,* vol. 104, no. 3, pp: 418–425, 2011.

[8] H. Golpira, & H. Danyali, "Reversible blind watermarking for medical images based on wavelet histogram shifting," in *Proceedings of IEEE International Symposium on Signal Processing and Information Technology*, pp. 31–36, 2010.

[9] M.S. Nambakhsh, A. Ahmadian, M. Ghavami, R.S. Dilmaghani, & S. Karimi-Fard, "A novel blind watermarking of ECG signals on medical images using EZW algorithm," in *Proceedings of the 28th IEEE-EMBS Annual International Conference*, pp. 3274–3277, 2009.

[10] R. Rodriguez-Colin, F. U. Claudia, & G. D. Trinidad-Blas, "Data hiding scheme for medical images," in *Proceedings of the 17th International Conference on Electronics, Communications and Computers*, CONIELECOMP '07, pp. 32–37, 2007.

[11] M. Fallahpour, D. Megias, & M. Ghanbari, "Reversible and high-capacity data hiding in medical images," *IET Image Process*., vol. 5, no. 2, pp: 190–197, 2011.

[12] G. Coatrieux, L. Lecornu, B. Sankur, & C. Roux, "A review of image watermarking applications in healthcare, Engineering in Medicine and Biology Society," EMBS '06, in *Proceedings of the 28th Annual International Conference of the IEEE*, pp. 4691–4694, 2006.

[13] K. A. Navas, & M. Sasikumar, "Survey of medical image watermarking algorithms," in *4th International Conference: Sciences of Electronic, Technologies of Information and Telecommunications*, TUNISIA, pp. 1–6, 2007.

[14] M. J. Tovée, "An Introduction to the Visual System," Second Edition, *Cambridge University Press*, 2008.

[15] M. A. Usman, G. B. Satrya, M. R. Usman, & S. Y. Shin, "Detection of small colon bleeding in wireless capsule endoscopy videos," *Computerized Medical Imaging and Graphics,* Vol. 54, pp: 16-26, 2016.

[16] D. A. Huffman, "A method for the construction of minimumredundancy codes," *Proceedings of the IRE*, vol. 40, no. 9, pp. 1098-1101, 1952.

[17] C. P. Wu, & C. C. Kuo, "Design of integrated multimedia compression and encryption systems," *IEEE Transactions on Multimedia*, vol. 7, no. 5 pp. 828-839, 2005.

[18] F. Özkaynak, A. B. Özer, & S. Yavuz, "Analysis of chaotic methods for compression and encryption processes in data communication," *in 20th Signal Processing and Communications Applications Conference* (SIU), pp. 1-4, Apr. 2012.

[19] G. Jakimoski, & K. P. Subbalakshmi, "Cryptanalysis of some multimedia encryption schemes," IEEE Transactions on Multimedia, vol. 10, no. 3, pp.330-338, Apr. 2008.

[20] K. K. Tseng, J. M. Jiang, J. S. Pan, L. L. Tang, C. Y. Hsu, & C. C. Chen, "Enhanced Huffman coding with encryption for wireless data broadcasting system," in *Computer, Consumer and Control (IS3C), 2012 International Symposium on*, pp. 622-625, Jun. 2012.

[21] Y. S. Jang, M. R. Usman, M. A. Usman, & S. Y. Shin, "Swapped Huffman tree coding application for low-power wide-area network (LPWAN)," In International Conference on Smart Green Technology in Electrical and Information Systems (ICSGTEIS), IEEE, pp: 53-58, Oct. 2016.