International Journal of Scientific Engineering and Applied Science (IJSEAS) – Volume-2, Issue-4,April 2016
ISSN: 2395-3470
www.ijseas.com

Evaluating Data Compression and Image Steganography Techniques for Optimized Embedding

Sanjay Bajpai[1], Kanak Saxena[2]
[1] Lakshmi Narain College of Technology-MCA, Bhopal, India
[2] Samrat Ashok Technological Institute, Vidisha, India

Abstract
We investigate the performance of state of the art data compression techniques and steganography methods for digital images proposed in the literature. Data compression and hiding methods are tested against a number of well-known steganographic methods for digital images that operate in spatial domains. Our experiments are performed and results are analyzed on large data set of digital colored images captured by our own canon camera and images available on the websites. Images are categorized on the basis of size, texture and types that include only jpeg, bmp and png types to determine their impending impact on steganographic methods. Text data to be embedded is classified and selected of varying capacities and domains to determine best suited data compression mechanisms and steganographic techniques. Our results designate the selection of compression and embedding techniques with respect to types of data.
Keywords: *Data compression, Distortion, Embedding techniques, Spatial domains.*

1. Introduction

Development in the science and technologies has facilitated in many ways but history is the proof that it has also degraded the moral values of most of the people. A new invention requires the endeavor of many scientists for the years but history reveals the facts that people do not hesitate in using it in wrong and unlawful acts. Often it is used to harm others and same case is with the computer technologies. Rapid development of computer and internet technologies has made the life easy and fast but at the same time, it has also enabled people to crack the security measures and peer into our private and secret information. Steganography has evolved as the most popular technique for securing the information and many methods have been proposed in the literature.

Security imposing methods like watermarking [1-2], finger printing and cryptography has its importance in its own way that varies from applications to applications. Many different methods have been proposed by various researchers for hiding text messages in the digital images but there is a lack of assessment on, that which type of steganographic method will be more suitable for hiding which type of text message. Some of the methods discussed in the literature are LSB substitution methods using variation of bits, modulus based LSB methods [3], Genetic Algorithm (GA) based LSB [4], Pixel Value Differencing (PVD) methods [5], embedding in the skin tone regions [6]. Some more methods that are reviewed and discussed in the literature are Bit Plane Complexity Segmentation (BPCS) method [7], Block Truncation Coding (BTC) method [8], location based embedding that is done on the occurrence of Most Frequent Pixels (MFP) and Second Most Frequent Pixels (SMFP) [9], run length approach [10], Content Based Image Embedding technique that segments homogeneous gray scale areas [11], seam carving methods that determines low energy and high energy pixels [12] and Pixel Indicator Technique (PIT) [13]. Some authors also incorporated data compression techniques to increase the embedding capacity and security level, such as, arithmetic coding [14] and Huffman coding [15] that improves the performance to a certain extent but it lacked the generalization of text messages. Both these techniques are analyzed to obtain the statistics which are used to train the classifier to set up the relationship between the message types and embedding and/or compression techniques. Ratnakirti Roy and et. al [16] considered many steganographic methods both in spatial and transform domain to perform their evaluation. They described the embedding mechanism of every algorithm under consideration but did not include the correlation between message types and embedding mechanism. Thus, there is a lack of study that

provides a correlation between the two. Our goal in the proposed work is to evaluate data hiding algorithms in digital colored images and data compression techniques and to correlate them with different types of messages. In this panorama, we would like to answer some questions like

1. What is the impact of size and texture of an image in selecting the text message?
2. What factors affect the selection of a data compression technique?
3. How does message size affect the selection of a data hiding algorithm?
4. What is the correlation between the embedding capacity and color distortion caused in the stego image?

These questions will be answered in section 3 with comparative analysis of results. Remainder of the paper is organized as follows. Section 2 includes the data set and its descriptions used in the experiments. Section 3 discusses embedding and compression techniques along with the nature of results obtained based on the varying inputs. Result derivatives are discussed in section 4. Finally conclusion and constraints are presented in section 5.

2. Data Set Descriptions

Efficiency and efficacy of any performance evaluation work depends on the data set selected for the experiments and one has to be more cautious in this regard to prove the work. Initially, we focused on forming the two data sets, namely, for the cover images and the text messages. Data set of images is again categorized into two sets, one set comprises of images in their original form as they are captured by the camera or available in the websites and second set comprises of those images that have been pre-processed to give unpredictable look and we gave the name to such images as mottled image [17]. We aimed to include images in the first data set that comprises of a variety of texture and sizes but mottled images are classified only on the basis of size. To prove the robustness of algorithms in certain cases, we created the environment to form special images and captured them by the camera. In the second data set, we included a variety of text messages that are categorized on the basis of size and domains. Having the variety in types, sizes and texture, we demonstrated the categorization in the form of tables.

Table 1 Original Image

File Formats	Size (in KB)			Texture	
	Small	Medium	Large	Low gradient	High gradient
JPEG	10–250	251 – 700	701–1200	Normal change in color	Drastic change in color
BMP	10–150	151 – 400	401–800		
PNG	10–100	101 – 300	301–600		

Table 2 Mottled Image

File Formats	Size (in KB)		
	Small	Medium	Large
JPEG	100 – 400	401 – 900	901 – 1500
BMP	50 – 300	301 – 700	701 – 1200
PNG	25 – 200	201 – 500	501 – 800

Table 3 Text Message Classification

Types	Size (in characters)			
	Very small	Small	Medium	large
General	4 – 10	11–400	401–10000	10000 – 10 lakhs
Domain based	–	100–1000	1001–20000	20001 – 15 lakhs

Table 1 contains the classification of day to day life images which are named as original images on the basis of size and texture. Here, we considered two types of textures namely low gradient and high gradient. Low gradient specifies those images, in which rate of change of color is low where as in high gradient, color changes drastically in the images. Table 2 contains the categorization of those images that have been pre-processed by masking their bits to reflect unpredictable behavior, named, mottled images. These images are classified only on the basis of size and texture does not play any role because of

their appearance that is discussed in section 3. Table 3 categorizes text messages which belong to general category and domain based category. Here, general means any text message conveying some meaning and domain based means that a message belongs to a particular literature area like text from computer science, medical science, accounting literature etc.

3. Experimental Setup

Steganographic methods are basically based on LSB substitution methods which are based on the constraint that cover image and stego image must be identical to eliminate the point of attack. Almost all the methods discussed in the literature like Pixel Value Differencing (PVD) methods [5], method based on Optimal Pixel Adjustment Procedure (OPAP) [18], methods proposed by Rosziati and et.al [19] and Mahmud Haasan and et. al [9] and many more are based on this approach but these are constrained to limited embedding capacity. In contrast to this, a method which is not based on the LSB substitution approach, discussed in [17], is also included for evaluation to give the horizon touching embedding capacity. An architectural framework is designed to test all these algorithms. Any of the platforms and programming languages could have been chosen but we implemented in Java because it facilitated the bit processing very efficiently. Two compression techniques are included in our study, namely, Arithmetic coding and Huffman coding. Data hiding algorithms included in the study to train the classifier are:

(a) SIS (Steganography Imaging System) proposed by Rosziati and et. al [19] converted the text file containing the secret message into the binary codes. They embedded two bits in each pixel implying that four pixels will be needed to embed one character completely.

(b) PVD methods suggest embedding in those portions of the image which are less susceptible to the HVS (human visual system). Many of the algorithms based on PVD methods can be referred in [5], [20], [21]. Empirical relations show that the pixels in the edge area are more suitable for embedding compared to the edge areas because the difference between the pixels in smooth area is less than that of edge area which ultimately leads to less color distortion.

(c) Mahmud Hasan and et. al [9] proposed the method, named *block processing mechanism*, in which cover image is divided into non-overlapping blocks of dimension 4×4. Most Frequent Pixels (MFPs) and Second Most Frequent Pixels (SMFPs) are identified and deleted from the occurrence and remaining pixels are used for embedding the secret message.

(d) A multi-key LSB substitution method proposed by Sanjay Bajpai and et.al [22] compartmentalizes pixels into its RGB components and segments the bits of a character. Applying permutation and combination on the LSBs of the components and bits of the character, text message is embedded into cover image.

(e) An innovative approach proposed by Sanjay Bajpai and et. al [17] generates mottled image by masking the bits of the cover image to embed text message in all the bits of the RGB components to enhance the embedding capacity that removes the constraint that cover image and stego image must be identical.

Many of the two data compression techniques are included in the study to increase the security level, which are:

(i) Arithmetic encoding: In this method, each character is encoded by assigning an upper bound and lower bound value, which lies in the interval of 0 and 1. Upper bound and lower bound of next character is calculated by using the upper bound and lower bound of previously encoded character. Detailed description of this method can be referred in [14]. As the number of character to be encoded increases, the upper bound and lower bound converges to same value and hence cannot be applied further. Results reveal that it can be applied for the very small secret messages whose length is less than 10.

(ii) Huffman encoding: This compression technique generates the unique binary code for each unique character in the text message. All the unique characters are represented in the form of binary tree with the help of which binary codes are generated. Its detailed process can be referred in

[15]. Length of the binary code depends on the frequency of the character occurring in the text message and it is inversely proportional to frequency. Results show that, if text message belongs to a particular domain like computer, medical or finance etc. then most of the words will repeat which ultimately leads to the higher frequency of most of the characters [23]. Thus, length of binary codes will be less compared to text message of general area and hence larger message can be embedded leading to higher embedding capacity.

3.1 Observed Facts

Some facts are observed while implementing the algorithms and analyzing the results, that if adopted, provide better selection of cover images and do not bind strictly to message and image types.

(i) Selection of image should not only be based on the size but its dimensions must also be checked. We checked about 500 images and observed that for the same dimension, size of the images varies. Some facts are mentioned in Table 4. Empirically, it is concluded that size of the images vary for the same dimension because of the colors of the image. Images having the vibrant colors are found to be of more size compared to others for the same dimension.

(ii) Resizing the image according to length of the text message reduces computation time and enables optimal embedding and transmission over the network.

(iii) Huffman encoding is more suitable for the text if it belongs to a particular domain. It not only increases the security level but also enhances the embedding when medium type images are used as cover images.

Table 4 Dimensions of Images

Image Type	Dimension (in pixels)	Size
Small image	450×338	57.2 KB to 87.6 KB
Medium image	1600×1200	355 KB to 727 KB
Large Image	2816×2112	0.8 MB to 1.47 MB

4. Result Derivatives

Inferences derived from the analysis of results are summarized in Table 5 which guides and makes the selection process easy. It only acts as the guiding element and options can be changed as per the availability and requirements. In certain cases, it has to be followed strictly such as, arithmetic coding can only be applied for very small messages. Similarly, Huffman coding does not imply much sense in case of small messages but advised to be used for medium and large messages. Furthermore, if mottled image is used as the cover image then it is advised not to use Huffman coding since it will unnecessarily increase the computational complexity and these types of images are themselves capable of hiding huge amount of text messages. Image in Fig. 1 is the original image as captured by the camera of size 1.81 MB and dimension 2816×2112 that has been resized to approximately match with the length of the text message as shown in Fig. 2 and its new size and dimension becomes 187 KB and 800×600 respectively. Image in Fig. 3 is the mottled image obtained by masking all of its bits.

Table 5 Classifier

Message Types	Compression Technique	Steganographic Methods	Image Types
Very small	Arithmetic coding	SIS, PVD	small
Small	–	SIS, PVD, Mahmud Hasan's process	small
Medium	Huffman coding	Multi-key embedding	medium
Large	Huffman coding	Multi-key embedding	large
Large	–	Mottled	large

Figure 1Original image captured by camera

Figure 2 Resized image to match size of text message

Figure 3 Mottled image

4. Conclusions

This paper evaluates some of the steganographic algorithms proposed in the literature and proposes a guiding element that facilitates in selecting the appropriate embedding method and compression technique. There are many data hiding algorithms for digital images but we considered only methods proposed by Rosziati and et. al [19], Mahmud Hasan and et. al [9] and PVD based methods for very small and small messages. Other methods can also be considered for this case, because despite of different embedding mechanisms, embedding capacity is almost the same. Other methods discussed in the paper [17], [22-23] are for medium and large messages. Much focus is emphasized on jpeg and png images as they occupy less space in the memory compared to bmp images and preferred for data transmission. This model requires thorough knowledge about the various steganographic techniques and basic operations that are performed on images like cropping, resizing etc. so that one can manipulate them to make fit in the model.

References

[1] C. I. Podilchuk, E. J. Delp, Digital watermarking: algorithms and applications, IEEE Signal Process. Mag. 18(4) (2001), pp. 33–46.

[2] J. J. K. O' Ruanaidh, W. J. Dowling, F. M. Boland, "Watermarking digital images for copyright protection", IEE Proceedings on Vision, Image and Signal Processing 143(4) (1996), pp. 250–256.

[3] V. Vijayalakshmi, G. Zayaraz, and V. Nagaraj, "A modulo based LSB steganography method", IEEE Conference on Control, Automation, Communication and Energy Conservation, August 2009, pp.1-4.

[4] A.M. Fard, M.R Akbarzadeh and A. F Varasteh. "A New Genetic Algorithm Approach for Secure JPEG Steganography," International Conference on Engineering of Intelligence Systems, 2006, pp 1-6.

[5] D. C. Wu, & W. H. Tsai, "A steganographic method for images by pixel-value differencing", *Pattern Recognition Letters, vol. 24, no. 9-10*, 2003, pp. 1613-1626.

[6] Anjali A. Shejul, Umesh L. Kulkarni, A Secure Skin Tone based Steganography Using Wavelet

Transform, International Journal of Computer Theory and Engineering, Vol.3, No.1, February 2011, pp. 16-22.

[7] J. Spaulding, H. Noda, M. N. Shirazi, E. Kawaguchi, BPCS steganography using EZW lossy compressed images, Pattern Recognition Letters 23(13) (2002). Pp.1579–1587.

[8] Chin-Chen Chang, Chih-Yang Linb, Yi-Hsuan Fan, Lossless data hiding for color images based on block truncation coding, The Journal of Pattern Recognition, Elsevier 41 (2008), pp. 2347–2357.

[9] Mahmud Hasan, Kamruddin Md. Nur, Tanzeem Bin Noor, A Novel Compressed Domain Technique of Reversible Steganography, International Journal of Advanced Research in Computer Science and Software Engineering ISSN: 2277 128X, March 2012, pp. 1-6.

[10] Chin-Chen Chang, Chih-Yang Lin, Yu-Zheng Wang, New image steganographic methods using run-length approach, International Journal of Information Sciences (176), Elsevier, 03 February, 2006, pp. 3393-3408.

[11] J. Kong, H. Jia, X. Li, Z. Qi, A novel content-based information hiding scheme, in: Proceedings of the International Conference on Computer Engineering and Technology, 22–24 January 2009, vol. 1, pp. 436–440.

[12] K. Thilagam, S. Karthikeyan, Optimized Image Resizing using Piecewise Seam Carving, International Journal of Computer Applications (0975 – 8887) Volume 42– No. 14, March 2012, pp. 24-30.

[13] Adnan Gutub and et al., "Pixel indicator high capacity technique for RGB image based Steganography", WoSPA 2008 – 5th IEEE International Workshop on Signal Processing and its Applications, University of Sharjah, Sharjah, U.A.E. 18 – 20 March 2008.

[14] Paul G. Howard, & Jeffrey Scott Vitter, "Practical Implementations of Arithmetic Coding", *International Conferences on Advances in Communication and Control (COMCON 3), British Columbia, Canada*, 16-18 October, 1991, pp. 1-34.

[15] Pushpa R. Suri and Madhu Goel, "Ternary Tree and Memory-Efficient Huffman Decoding Algorithm", IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 1, ISSN (Online): 1694-0814, pp. 483-489, January 2011.

[16] Ratankirti Roy, Suvamoy Changder, Anirban Sarkar, Narayan C Debnath, Evaluating Image Steganography Techniques: Future Research Challenges, 978-1-4673-2088-7/13/$31.00 ©2013 IEEE, 2013, pp. 309-314.

[17] Sanjay Bajpai, Kanak Saxena, A High end Capacity in Digital Image Steganography: Empowering Security by Mottling through Morphing, International Journal of Scientific and Engineering Research (IJSER), Volume 5 Issue 2, ISSN 2229-5518, February 2014, pp. 1081-1086.

[18] Chi-Kwong Chan, & L.M. Cheng, "Improved hiding data in images by optimal moderately significant-bit replacement", *IEE Electron Lett. 37 (16)*, 2001, pp. 1017-1018.

[19] Rosziati Ibrahim and Teoh Suk Kuan, Steganography algorithm to hide secret message inside an image, Journal of Computer Technology and Application 2 (2011), 25 February, 2011, pp. 102-108.

[20] Xin Liao, Qiao-yan Wen, Jie Zhang, A steganographic method for digital images with four-pixel differencing and modified LSB substitution, Journal of Vis. Commun. Image Elsevier R(22) (2011), pp. 1-8.

[21] C. H. Yang, C. Y. Weng, A steganographic method for digital images by multi-pixel differencing, in: Proceedings of International Computer Symposium, Taipei, Taiwan, R.O.C., 2006, pp. 831–836.

[22] Sanjay Bajpai, Kanak Saxena, **"**Enhancing Embedding Capacity by Compartmentalizing Pixels using LSB Techniques in Steganography", *International Journal of Computers and Applications*, ACTA Press, paper-id 202-3762, submitted – 03 April, 2013.

[23] Sanjay Bajpai, Kanak Saxena, **"**Enhancement of Security and Embedding Capacity through Huffman Coding in Steganography", *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Volume 2, Issue 4, ISSN 2278-6856*, July – August, 2013, pp. 73-78.

Sanjay Bajpai received his MCA degree from Rani

Durgawati University, Jabalpur in Jan, 1999, M. Tech. (s/w engg.) degree from RGPV, Bhopal in June, 2011 and Ph.D. degree in Computer Science from Barkatullah University, Bhopal in February, 2016. Currently, he is working as Professor and vice Principal in LNCT-MCA, Bhopal. He is having 15 publications in International/National Journals/ Conferences. His research area is data security and image processing and has published one question bank with answers for the subject Artificial Intelligence & Applications.

Kanak Saxena has received MCA, M. Tech. degree and Ph. D. degree in Computer Science from DAVV, Indore. She is working as Professor and Head (Computer Applications Deptt.) SATI, Vidisha. She is having 96 publications in International/National Journals/ Conferences. She is Ph.D examiner of various universities and has guided 9 Ph.D. students. Her area of interest is Network security, Cloud computing and Image processing.