A Novel Digital Image Steganalysis Approach for Investigation

Nanhay Singh
Ambedkar Institute of Advanced Communication Technologies and Research, Delhi, India.

Bhoopesh Singh Bhati
Ambedkar Institute of Advanced Communication Technologies and Research, Delhi, India.

R. S. Raw
Ambedkar Institute of Advanced Communication Technologies and Research, Delhi, India.

ABSTRACT

In the electronic world, one of the most appropriate "hosts" for steganography are digital images. Digital photographs are a commonly shared, sent, and exchanged throughout the Internet in the form of email attachments or web postings. However, current steganographic software available on the market has poor support for high-capacity image steganography. Even worse, some steganographic software actually distorts or degrades the appearance of cover images and therefore exposes the steganographic transformation the image has undergone. In this paper, we propose an algorithm for investigate digital image by steganalysis which is extended version of Modified Kekre Algorithm.

Keywords

Steganalysis, Digital Forensic, Digital Image Steganalysis, Extended Version of MKA

1. INTRODUCTION

Steganography is a superannuated artistry of conveying messages in a secret way that only the receiver knows the existence of message. So, fundamental assert for a steganographic technique is imperceptibility; this means that the embedded messages should not be discernible to the human eye. This field is expected to flourish. Steganography might also behove limited under regulations, since governments already described that criminals take advantage these techniques to communicate. More confine on the use of privacy-protecting technologies are not very unlikely, especially in this period of time with great anxiety of terrorist and other attacks.

A complete steganographic transaction essentially consists of three separate processes: encoding, transmission, and decoding. Encoding consists of all the tasks performed in embedding hidden data in an image. Transmission refers to the transfer of the stego image from a sender to a recipient. Decoding involves the extraction of the hidden data from the stego image.

High Capacity has major concern in image steganography. The proposed method in image steganography is used to improve the capacity of investigation for hidden data into hosted signal without causing any statistically significant modification. Many novel data hiding methods based on Least Significant Bits (LSB) and Pixel Value Differencing (PVD) to increase the hiding capacity have been proposed with imperceptible quality. One of the above methods we have improved the Modified Kekre's Algorithm (MKA) [3]. This is based on LSB technique. The improved scheme increases the embedding capacity while retaining the quality of stego image (carrying hidden data) as good as MKA [1]. Experimental results show that the improved scheme outperform the original comparative scheme especially in capacity of hidden data-bits where higher intensity of the pixel decide the number of bits to embed into the cover-image [1]. According to figures 1 and 2, lower intensity pixel cannot distort the visual quality of pixel and it can also store higher number of bits. Our improved version shows that we are efficiently utilizing the maintain matrix (it maintain the position of pixel where 5 LSB are used to embed the data).

The rest of this paper is arranged as follows: Section 2 gives an overview about the Background and related work in the area Forensic Investigations by steganalysis for digital image. Section 3 the detail of the Modified Kekre's Algorithm. Section 4 the detail of the problem definition. Section 5 proposed algorithm for the better response by Modified Kekre's Algorithm. Section 6 experimental evolution of our propose algorithm with an example. Finally, some conclusion and prospect are put forward in Section 7.

2. BACKGROUND AND RELATED WORK

Many researchers have looked for the way to represent the steganography and future of steganography. Some of them have said that Steganalysis is the technique for digital forensic investigation. In this research, authors have analyzed 200 images taken from 30 different Croatian and B&H websites. The American National Standard Code for Information Interchange is a standard for presenting of the text and the characters in a digital world. ASCII uses 7 bytes to present any character [8].

Kekre et al. [2] proposed for gray scale images where two consecutive pixel pairs are formed and embedding capacity is decided based on difference in their values. In this paper author extend PVD to 24 bit color images and also a new improved version of Least Significant Bit (LSB) method (Modified Kekre's Algorithm) which is also applied on 24 bit color images after that they proposed new technique for image retrieval using the color-texture features extracted from images based on vector quantization with Kekre's fast codebook generation is proposed. This gives better discrimination capability for CBIR. Here the database image is divided into 2x2 pixel windows to obtain 12 color descriptors per window (Red, Green and Blue per pixel) to form a vector. Collection of all such vectors is a training set. Then the Kekre's Fast Codebook Generation (KFCG) is applied on this set to get 16 code vectors. The Walsh transform is applied on each column of the codebook, followed by Kekre's transform applied on each row of the Walsh transformed codebook. This transform vector then is used as the image signature (feature vector) for image retrieval. The method takes lesser computations as compared to conventional Walsh applied on complete image. The method gives the color-texture features of the image database at reduced feature set size. Proposed method gives better precision and recall as compared to full Walsh based CBIR.

Proposed method avoids resizing of images which is required for any transform based feature extraction method.

Marvel et al. [5] propose a simple means of storing one bit per block in the quantized JPEG coefficients. Although this bears some surface resemblance to the work given here, the embedded data are stored in the JPEG coefficients themselves, and it is required that the receiver have them in order to extract the embedded data. In the method presented here, the data are encoded in the spatial domain, albeit through manipulation of the frequency domain, and the receiver must have the spatial domain realization of the image in order to extract the embedded data. In our baseline system, the receiver must also generate the topologically nearby spatial blocks in order to determine whether the block in question actually encodes data or is unusable. In follow-on work, this requirement is eliminated.

Recently there has been a great deal of work in investigation by steganalysis. The work is most closely related to the steganography that we study here. Here, describes a steganography based algorithm for digital image to identify the hidden context.

3. MODIFIED KEKRE'S ALGORITHM

Modified Kekre's Algorithm (MKA) [3] is based on Least Significant Bit (LSB) method. MKA is applied on 24 bit Read Green Blue (RGB) color image. It uses up to five LSB's of a pixel to embed the data. The intensity of the pixel value decides the number of LSB's to embed and control the error. MKA uses 8 bit secret key to perform XOR operation to all the bytes of message to achieve security. The embedding algorithm maintains a matrix of pixels where 5 bits of message are used to embed and this matrix is required extracting the hidden message from stegoimage. Following checking process decides the number of bits to embed into pixel. The term "MSB" is used as Most Significant Bit of cover-image pixel, and "Message Bit" represents a bit of message to hide. Don't care bits are represented with 'x'. In the given pseudo code of MKA algorithm which shows that up to 5 LSB of pixel can be used to embed the bits of message data depending to the intensity of the pixel value. The above procedure also has been mapped into table. Where 4 MSB are 1111 another 4 LSB xxxx are don't care (whatever bits it contained in 4 LSB's) and the data bit (want to embed) is 1 then utilize 5 LSB of pixel and mark maintain matrix pixel position to identify that this pixel contains 5 bits of data. If data bit is not 1 then we use only 4 LSB of pixel to embed data bit. Same procedure will be run to extract the data bits of message using estimated matrix because it keeps the track of the pixel position where 5 LSB bits are utilized. At the end 8 bit secret key with XOR operations is applied on the extracted message to regenerate original message which was previously embedded [7].

Table 1: Data Embedding Based On Pixel Bits

S.No.	MSB's of Pixel	Matrix Entry	If Data Bit is	Utilize Bit/Bits
1	1111 xxxx	1	1	5
2	1111 xxxx	-	0	4
3	1110 xxxx	1	0	5
4	1110 xxxx	-	1	3
5	1100 xxxx	-	x	2
6	1000 xxxx	-	x	1

Where **x:** Don't care bit

4. PROBLEM DEFINITION

In this paper, we have improved MKA algorithm with respect to two major aspects. First lower intensity pixel have also been used for data hiding and secondly maximum utilization of matrix which keeps the track of pixel where 5 LSB are used.

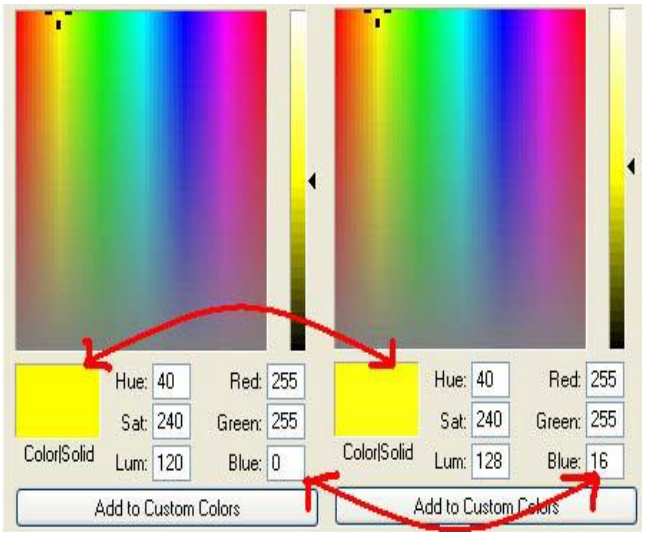

Figure 1: Extreme Modified 4 LSB's for One Channel

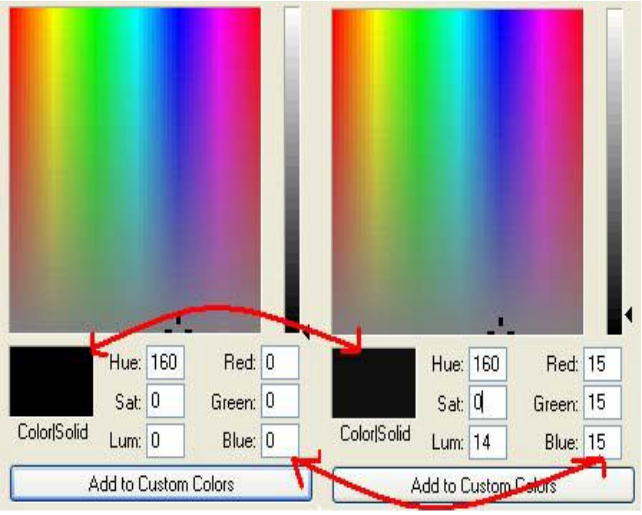

Figure 2: Extreme Modified 4 LSB's for Three Channels

So according to the figure-1, three pixels (Red = 255, Green = 255 and Blue = 0) are generating Yellow color in left part. If we change (Blue = 16) 4 LSB's of Blue pixel will generate Yellow color in right part of the figure-1. If we compare both left and right part, Yellow color almost has the same color and also same visual quality. Same scenario shown in figure-2 where if we modified the 4 LSB's of all pixels (RGB) then resulting color seems to be the same as figure-2 of left part. Our idea is that change in lower intensity pixel value has less visual degradation quality effects. If pixel intensity is less than 16, it can be modified into 0 to 15 ranges, which will not degrade the visual quality of that pixel. So we can also embed up to 4 bits into that pixel [2] [6].

5. PROPOSED ALGORITHM

In our proposed algorithm by apply 8 bit confidential data with XOR operation on all bytes of message to modify its originality of message as in MKA [4] and also maintaining a matrix for those pixels which embed 5, 3 and 2 LSB's of data. Following verification process decides to embed the data bits to cover image.

Algorithm: *Steganorensic*

Step 1. If 4 Most Significant Bit equals 1111 then

Step 2. If Message Bit equals 1 then utilize 5 Least Significant Bit with data & mark Else utilize 4 Least Significant Bit with data

Step 3. Else

Step 4. If 3 Most Significant Bit equals 111 then

Step 5. If Message Bit equals 0 then utilize 5 Least Significant Bit with data & mark Else utilize 3 Least Significant Bit with data

Step 6. Else

Step 7. If 3 Most Significant Bit equals 110 then

Step 8. If Message Bit equals 0 then mark with 1, move to next Message Bit & embed next 2 Message bits to 2 Least Significant Bit's. (utilized 3 Bits) Else utilize 2 Least Significant Bit with data

Step 9. Else

Step 10. If 2 Most Significant Bit equals 10 then

Step 11. If Message Bit equals 0 then mark the matrix with 1, move to next Message Bit & embed next 1 Message bit to 1 Least Significant Bit's. (utilized 2 Bits) Else utilize 1 Least Significant Bit with data

Step 12. Else

Step 13. If 2 Most Significant Bit equals 01 then

Step 14. If Message Bit = 0 then mark the matrix with 1, move to next MSG Bit & embed next 1 Message bit to 1 Least Significant Bit's. (utilized 2 Bits) Else utilize 1 Least Significant Bit with data

Step 15. Else

Step 16. If 3 Most Significant Bit equals 001 then

Step 17. If Message Bit equals 0 then mark the matrix with 1, move to next Message Bit & embed next 1 Message bit to 1 Least Significant Bit's. (utilized 2 Bits) Else utilize 1 Least Significant Bit with data

Step 18. Else

Step 19. If 4 Most Significant Bit equals 0001 then

Step 20. If Message Bit equals 0 then mark the matrix with 1, move to next Message Bit & embed next 2 Message bits to 2 Least Significant Bit's. (utilized 3 Bits) Else utilize 2 Least Significant Bit with data

Step 21. Else

Step 22. If 4 Most Significant Bit equals 0000 then

Step 23. If Message Bit equals 0 then utilize 5 Least Significant Bit with data & mark Else utilize 4 Least Significant Bit with data

In the above proposed algorithm, we have noticed that it has stored 2 bits when marking the entry in matrix. One bit in matrix and other is in pixel itself. One bit of data we are discarding and mark the matrix 1. Basically we are discarding a bit which already exists into a pixel then store another bit in 1st LSB of a pixel. The same case will apply when we are embedding 3 or 2 bits of data with mark matrix.

6. EXPERIMENTAL EVOLUTION

In our experiments, we have an example 10xx xxxx pixel value and data bits are 01 for embedding. After embedding value will be 10xx xxx1, first 0 will be replaced with 10xx xxxx pixel 7th bit of 0. Against of it mark the matrix at this position of pixel and next data bit 1 is replaced with 1st pixel LSB bit as 10xx xxx1. This also shows that we have used the low intensity based pixel with up to 5 LSB of data embedding and also comprehensive use of matrix. The above procedure has also been mapped into table-2. Same procedure will be run to extract the data bits of message using estimated or maintained matrix. Maintained matrix keeps the track of the pixel position where 2, 3 and 5 bits are utilized. At the end 8 bit secret key with XOR operation is applied on the extracted message to regenerate original message which was embedded [4] [6].

Table 2: Proposed Data Embedding Based On Pixel Bits

S. No.	MSB's of Pixel	Matrix Entry	If Data Bit is	Utilize Bit/Bits
1	1111xxxx	1	1	5
2	1111xxxx	-	0	4
3	1111xxxx	1	0	5
4	1111xxxx	-	1	3
5	1111xxxx	1	0	3
6	1111xxxx	-	1	2
7	1111xxxx	1	0	2
8	1111xxxx	-	1	1
9	1111xxxx	1	0	2
10	1111xxxx	-	1	1
11	1111xxxx	1	0	2
12	1111xxxx	-	1	1
13	1111xxxx	1	0	3
14	1111xxxx	-	1	2
15	1111xxxx	1	0	5
16	1111xxxx	-	1	4

Where x: Don't care bit

In this example the process is illustrated table 2 that contain data embedding based on pixel bits is presented.

7. CONCLUSION AND FUTURE WORK

The major contributions of this paper are digital image steganalysis algorithm for forensic investigation. The main advantage of our proposed algorithm is that we can use all the bytes of cover image to hide the data bits. Our proposed algorithm has very high analyzing capacity of data hiding as compare to MKA [1, 3]. Our grand goal is to develop algorithm that can be done at digital image, while respecting their investigation policies. In this paper, we proposed an algorithm for investigation on digital image steganalysis.

In future we aim to improve our algorithm and implement it in real dataset. Additionally we plan to investigate how to quality of published images will be improved.

8. REFERENCES

[1] H. B. Kekre, Archana Athawale and Pallavi N.Halarnkar 2009, "Polynomial Transformation To Improve Capacity Of Cover Image For Information Hiding In Multiple LSBs", International Journal of Engineering Research and Industrial Applications (IJERIA), Ascent Publications, Volume II, March 2009, Pune.

[2] Chen, W. J., Chang, C. C. and Le, T. H. N. 2010, "High Payload Steganography Mechanism Using Hybrid Edge Detector," Expert Systems with Applications (ESWA 2010), vol. 37, no. pp. 3292-3301, 4th April 2010.

[3] H. B. Kekre, A. Athawale and P. N. Halarnkar 2009, "Performance Evaluation of Pixel Value Differencing and Kekre's Modified Algorithm for Information Hiding in Images". ACM International Conference on Advances in Computing, Communication and Control (ICAC3).

[4] Hamid, A. M., M. L. M. Kiah, et al. 2009 "Novel Approach for High Secure and High Rate Data Hidden in the Image Using Image Texture Analysis." International Journal of Engineering and Technology (IJET).

[5] L.M. Marvel, G.W. Hartwig, and C. Boncelet 2000, Compression-compatible fragile and semi-fragile tamper detection. In SPIE EI Photonics West, pages 131{139, San Jose, CA.

[6] Hussain, M. 2010, "Pixel intensity based high capacity data embedding method" International Conference on Information and Emerging Technologies (ICIET), Pp.1 - 5.

[7] H. B. Kekre, Archana Athawale, Tanuja K. Sarode, Kalpana Sagvekar 2010, "Mixing Codebooks of LBG, KPE and KFCG Algorithms to Increase Capacity of Information Hiding", International Journal of Computer Applications, Number 3 - Article 4.

[8] N. M. Nasrabadi and R. King 1988, "Image coding using vector quantization: A review," IEEE Trans. Commun., vol. 36, no. 8, pp. 957–971, Aug. 1988.