Data Hiding in Digital Images: A Steganographic Paradigm

Synopsis of the Thesis to be submitted for the Partial Fulfillment of the Requirements for the Award of the degree of

Master of Technology

In

Computer Science and Engineering

Submitted by :
Piyush Goel(03CS3003)

Under the guidance of
Prof. Jayanta Mukhopadhyay

Department of Computer Science and Engineering,
Indian Institute of Technology Kharagpur
May, 2008

Abstract

In this work, we studied the steganographic paradigm of data hiding in digital images. Two approaches prevalent in current steganographic research were studied. The first approach is based on the concept of Statistical Preservation. The purpose of this approach is to restore the image statistics which get modified during embedding and are generally exploited by the steganalytic attacks. Several algorithms based on this approach exist in literature and the strengths and limitations of these algorithms are discussed. Two new algorithms which preserve the first order statistics of a cover image during embedding have been proposed. The first algorithm embeds data into a cover signal while inherently preserving the first order statistics of the image. The second algorithm makes an explicit attempt to restore the first order statistics of an image after completing the embedding operation on the image. The proposed algorithms are successful in breaking the targeted attacks based on the first order statistics of an image while overcoming the limitations of some of the existing schemes.

The second approach covered in this thesis is aimed towards defeating the blind steganalysis attacks especially the attacks based on the concept of Self-Calibration. We study some of the existing algorithms proposed in literature for breaking the calibration based blind attacks and analyze their strengths and weaknesses. A new framework called Spatial Desynchronization is proposed which can be used for resisting the calibration based attacks. This framework has been extended to a new steganographic algorithm called “Spatially Desynchronized Steganographic Algorithm”. It is shown experimentally that the proposed algorithm can provide higher payloads as well as higher security against calibration based attacks than the existing schemes.

1. Introduction

Steganography is the art of hiding information imperceptibly in a cover medium. The word "*Steganography*" is of Greek origin and means *"covered or hidden writing"*. The main aim in steganography is to hide the very existence of the message in the cover medium. Steganography and cryptography are counter parts in digital security the obvious advantage of steganography over cryptography is that messages do not attract attention to themselves, to messengers, or to recipients. Also, the last decade has seen an exponential growth in the use of multimedia data over the Internet. These include Digital Images, Audio and Video files. This rise of digital content on the internet has further accelerated the research effort devoted to steganography. The various applications of steganography include secure military communications, multimedia watermarking and fingerprinting applications for authentication purposed to curb the problem of digital piracy. Although these are not perfect applications of steganography, many steganographic algorithms can be employed for these purposes as well.

In this thesis we have explored two different approaches to steganography. The first approach as explained in next section is based on statistical preservation which preserves the statistics of an image which get distorted due to embedding and can be exploited by targeted attacks. This part of the work is explained in section 2. The second approach is based on preventing calibration based blind attacks which try to classify an image as stego or cover by trying to estimate the cover image statistics from the stego image by a process termed as Self – Calibration. This work has been discussed in section 3.

2. Statistical Preservation

Statistical undetectability is one of the main aspects of a steganographic algorithm. To maintain statistical undetectability, the steganographic techniques are designed with the aim of minimizing the artifacts introduced in the cover signal by the embedding technique. The main emphasis is on minimizing the noise added by embedding while increasing the payload. This is an important consideration in the design embedding algorithms, since the noise added effects the statistical properties of a medium. For a given medium, the steganographic algorithm which makes fewer embedding changes or adds less additive noise

will be less detectable as compared to an algorithm which makes relatively more changes or adds higher additive noise.

From the point of view of the steganalyst, the steganalytic attacks try to examine a signal and look for statistics which get distorted due to embedding. These statistics range from marginal statistics of first and second order in case of targeted attacks to extremely high order statistics (up to 9^{th} order) in the case of blind steganalytic techniques which use machine learning techniques for estimating a model of the cover image from these high order statistics and reports an image to be containing steganographic embedding if it does not conforms to this model. *So, in order to defeat the steganalytic attacks, there has been a shift from the above mentioned data hiding paradigm. Algorithms have been proposed which try to restore the statistics which get distorted during the embedding procedure and which may be used for steganalysis.*

In recent literature few algorithms have been proposed where marginal statistics are preserved for achieving more security. Solanki et al [1, 2] have proposed an algorithm where they have compensated the first order statistics (Image histogram) after the steganographic embedding. The compensation algorithm is based on a theorem proved by Tschoppe[4] which tries to convert one vector x into another vector y while satisfying a Minimum Mean Square Error (MMSE) criterion. This algorithm suffers from the following shortcomings:

- The algorithm assumes the cover image to be a Gaussian cover. The algorithm does not give good results for non-Gaussian cover images.
- The algorithm has been tried specifically for Quantization Index Modulation algorithm [10]. The algorithm is not giving good results for some well known steganographic embedding such as Least Significant Bit (LSB) Replacement, LSB matching etc.
- They have ignored low probability image regions for embedding due to erratic behavior in low probability tail.

In this work we propose a simple but effective embedding scheme called “Pixel Swapping” which inherently preserves the first order statistics of the cover image. Along with that a new scheme for statistical restoration has been proposed which overcomes the

shortcomings of the scheme given in [3]. The details of the proposed schemes are given in the next subsection.

2.1 Embedding by Pixel Swapping:

In order to maintain the first order statistics of the cover image, we have proposed a scheme which inherently preserves the first order properties of the signal. The scheme is outlined below:-

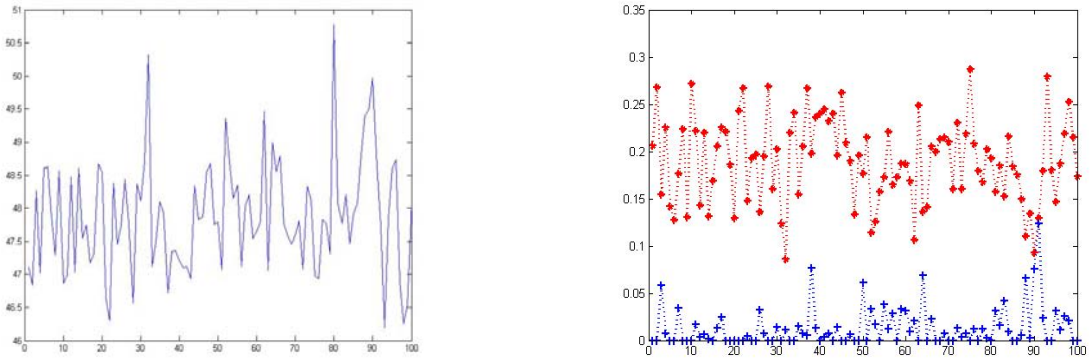
Algorithm Pixel Swap

- *Randomly select 2 pixels x1 and x2 from the cover image using a pseudo –random sequence.*
- *If the two pixels lie within a specified distance α (α=2 or 3 generally), they are suitable for embedding, otherwise generate another set of pixels.*
- *Pick up the message bit. If the message bit is zero (or one), check if x1 > x2 otherwise swap x1 and x2. Do the reverse operation for the message bit one (zero)*
- *For decoding, select the pixels using the same pseudo-random sequence. Check if the 2 pixels are within the pre-specified range α. If x1>x2, the message bit is zero (one) otherwise the message bit is one (zero).*

This scheme preserves the first order statistic (histogram) inherently without applying separate restoration process. This scheme also does not add any visual distortion to the image since the threshold used for swapping of pixels is kept considerably small (α <=5) which only affects the least significant bit planes of an image. To measure the distortion introduced by the embedding in the cover image, the Peak Signal to Noise Ratio (PSNR) after embedding was observed for one hundred images. It was found that the PSNR is constantly above 46 dB as seen in Fig 1(a). This scheme is able to resist the targeted attacks based on the first order statistic proposed in [6, 7] as shown in Fig 1(b). It should be noted that the attacks proposed in [6] and [7] are similar and hence we tested the performance of Pixel Swapping against the attack proposed in [6] only with the motivation that if it can resist the attack given in [6], then it would be able to resist the attack proposed in [7] as well.

One thing should be noted that the proposed scheme cannot ensure high embedding rates since the cover signal might have high spatial frequency thus reducing the number of pixel pairs satisfying the embedding condition. Another important point to be noted about the proposed scheme is that though it preserves first order statistics of the cover signal, it does

alter the second and higher order statistics of the cover. Hence this scheme can be detected using the Blind Steganalytic methods described in [8]. These attacks use very high order statistics of the cover signal and hence these very sensitive to even slight changes in the cover signal. Even a small amount of distortion can be detected using these blind schemes. In the next section a new statistical restoration technique has been proposed which can be used for preserving the histogram (first order statistics) of the cover signal after embedding.



(a) (b)

Fig1. (a) PSNR observed for 100 images after Pixel Swapping embedding with maximum embedding rate. (b) Performance of Pixel Swapping against Sample Pair Attack proposed in [6].

2.2 New Statistical Restoration Scheme:

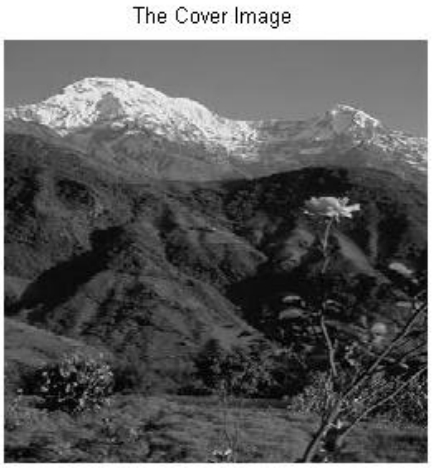
In this section we introduce a new scheme for statistical restoration which can be used for repairing the artifacts which get introduced in the cover signal. This scheme tries to overcome the shortcomings of the Solanki et.al's scheme [1, 2]. The proposed scheme is based on the idea of pixel swapping introduced above. The cover pixels are categorized into two streams, one is for embedding and another is for restoration. At the time of embedding value of a pixel (say α, from embedding stream) is changed to β. Now the idea is to find a pixel with value β in compensation stream and change it to α. But problem with this formulation is that at the time of embedding some pixels with value β may get changed to α. So there is no need of compensation at all. To overcome this problem, at the time of embedding we maintain a record of the pairs of pixel values which get changed into one another. So, after embedding we can get an exact count of the number of pixels which have to be compensated in order to maintain the first order statistics of the cover image.

Next important point to be noted is how much distortion is added to the cover due to the compensation procedure. This distortion is somewhat dependent on how much maximum change is made per embedding by the steganographic algorithm. In our experiments, we have used the ± 1 embedding. So the absolute distortion per pixel due to embedding is at most 1. So during the compensation step, the bin value is repaired using modification of immediate neighbors (immediate left or immediate right) of that bin, satisfying lowest mean square error due to compensation methods. The proposed method of compensation can easily be extended for ±k embedding or for that matter any kind of embedding procedure either in the spatial or the transform domain. But the amount of noise added due to compensation will increase with the increase in the noise added during the embedding step.

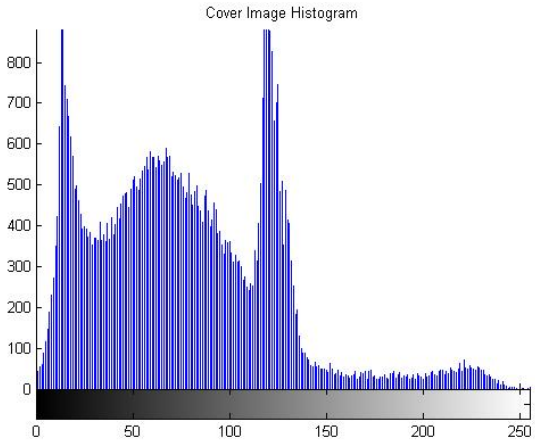
With regards to the efficiency of restoration process, the proposed restoration scheme is compared against the existing scheme proposed by [1, 2]. It was observed that the proposed scheme is able to achieve better histogram restoration especially for non-Gaussian cover distributions as can be seen in Fig. 2. The cover image has a non-Gaussian distribution (Fig. 2(b)) and the difference histogram between the cover and stego image before compensation is shown in Fig. 2(c). Fig. 2(d) shows the difference histogram after compensation using scheme proposed by Solanki. et. al. [1] and Fig. 2(e) shows the difference histogram after compensation using the proposed scheme. It can be seen that the proposed scheme provides better restoration than the Solanki.et.al scheme. The detailed analysis and proof of minimum noise of the proposed algorithm will be provided in the final thesis.

2.3 Security Analysis

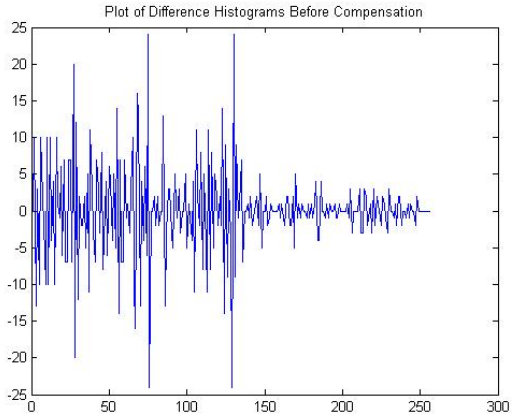
To evaluate the security of the two proposed schemes Pixel Swapping and Statistical Restoration, we have used one targeted attack called Sample Pair Analysis [6] and one blind attack called Wavelet Moment Analyzer (WAM) proposed by Fridrich et. al. [8]. The details of these attacks can be found in the [6, 8] and the performance of the proposed schemes will be covered in detail in the thesis. It was observed that both the proposed schemes were able to resist Sample Pair attacks successfully but did not give promising results against the WAM This was attributed to the fact that although we were able to restore the marginal statistics, but we introduced additional noise to the cover in this process. The complete results will be provided in the final thesis.



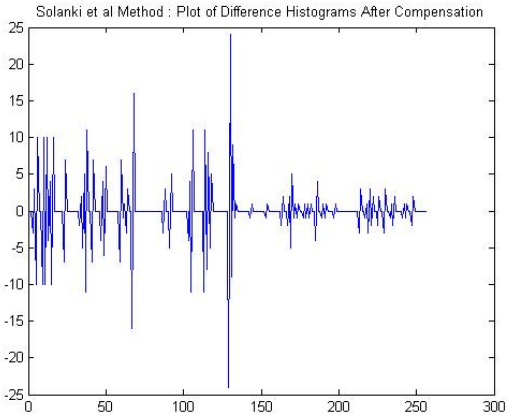
(a) Cover Image



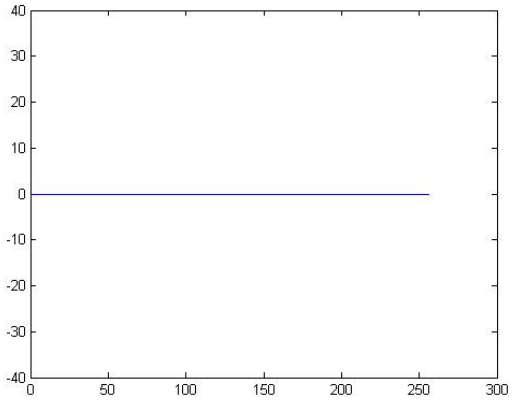
(b) Cover Image Histogram



(c) Difference Histogram Before Compensation



(d) Difference Histogram after compensation using Solanki's scheme



(e) Difference Histogram after compensation using Proposed Scheme

Fig.2. Performance Comparison of the proposed restoration scheme and the Solanki. et. al's scheme

3. Resisting Blind Attacks

In this part of the thesis, we studied some of the existing blind attacks in steganalysis literature and analyze their strengths and limitations. Blind attacks often try to estimate the cover image statistics from the stego image by nullifying the changes made in the cover signal by the embedding process. Then powerful machine learning techniques are used to classify an image as cover or stego based on these statistical features.

We specifically studied the "self-calibration" based blind attacks proposed in [5, 9]. These attacks are specifically designed to detect the presence of steganography in JPEG compressed images and are some of the most successful attacks for breaking JPEG steganography. We analyzed the performance of these attacks using a *Statistical Hypothesis Testing framework*. The tests were performed with testing two main goals:

1) Test the sensitivity of the features used in the calibration attacks.
2) Test the effectiveness of the self-calibration process.

The problem was reduced to testing whether two given samples are drawn from same population or not. The Rank Sum (Mann-Whitney Wilkoxon Test) provided by statistical toolbox of MATLAB v7.1 was used for conducting the test and the *p-values* were observed. The hypothesis testing was performed for features extracted for the attacks given in [5, 9]. It should be pointed out that the two attacks extract features in different ways and hence the problems of hypothesis testing were formulated in separately. For the first test of testing the feature sensitivity, it was observed that as the embedding rate was increased the *p-values* obtained decreased indicating that the two populations were statistically moving further away with respect to the features obtained by the two attacks.

We also studied some of the existing approaches proposed for defeating the calibration attacks [11, 12] and analyzed their strengths and weaknesses. We propose a new framework called Spatial Desynchronization which tries to disturb the prediction process of the blind attack. The key feature of this framework is that it embeds data in a spatially desynchronized version of the cover image. At the time of steganalysis, the steganalyst uses the portion of the image where data is hidden for predicting the cover image statistics and hence is not able to differentiate between the cover image and the stego image. This framework is then extended to a new steganographic scheme called *Spatially Desynchronized*

Steganographic Algorithm (SDSA). The main advantage of the proposed algorithm over some of the existing schemes is that it can achieve higher embedding rates than the existing schemes while maintaining same levels of security. The proposed algorithm is analyzed in the light of the statistical framework proposed above. It is shown that proposed algorithm produce stego images which are statistically closer to the cover image population. The robustness of the proposed algorithm is also tested against the calibration attacks and it is shown that the proposed algorithm could embed more payload into the cover signal while providing better levels of security. To check the performance of the *SDSA* algorithm against blind attacks not based on self-calibration, the performance was also tested against the attack proposed in [13]. The detailed results and analysis of this work will be provided in the final thesis.

Bibliography:

1) K. Solanki, K. Sullivan, U. Madhow, B. S. Manjunath, and S. Chandrasekaran, "Provably secure steganography: Achieving zero K-L divergence using statistical restoration", in Proc. *IEEE International Conference on Image Processing (ICIP06), Atlanta, GA, USA*, Oct. 2006.

2) K. Solanki, K. Sullivan, U. Madhow, B. S. Manjunath, and S. Chandrasekaran, "Statistical Restoration for Robust and Secure Steganography" in Proc. *IEEE International Conference on Image Processing, Genova, Italy*, Sep. 2005.

3) J. Fridrich, T. Pevny, and J. Kodovsky, "Statistically Undetectable JPEG Steganography: Dead Ends, Challenges, and Opportunities", in Proc. *ACM Multimedia and Security Workshop, Dallas, TX*, September 20-21, pp. 3-14, 2007

4) R Tzschoppe, R. Bäuml and J J. Eggers, "*Histogram Modifications with Minimum MSE Distortion*", Technical Report, December 18, 2001, Erlangen, Germany.

5) J. Fridrich, and T. Pevny, “Merging Markov and DCT Features for Multi-Class JPEG Steganalysis”, in Proc. *SPIE Electronic Imaging, Photonics West*, January 2007, pp. 03-04

6) S. Dumitrescu, X. Wu, and N. Memon, “On steganalysis of random lsb embedding in continuous-tone images,” in Proc. *IEEE International Conference on Image Processing,* Rochester, New York., September 2002.

7) J. Fridrich, M. Goljan and R. Dui,”Reliable Detection of LSB steganography in Color and Grayscale Images,” in Proc. of the *ACM Workshop on Multimedia and Security, Ottawa, CA*, October 5, 2001, pp. 27-30.

8) J. Fridrich, M. Goljan and T. Holotyak “New Blind Steganalysis and its Implications”, with, in Proc. *SPIE Electronic Imaging*, Photonics West, January 2006

9) J. Fridrich, “Feature-Based Steganalysis for JPEG Images and its Implications for Future Design of Steganographic Schemes”, in Proc. *6th Information Hiding Workshop*, *Toronto , Canada*, pp. 67-81, May 2004.

10) K. Solanki, N. Jacobsen, U. Madhow, B.S. Manjunath, and S. Chandrasekaran, *“*Robust image-adaptive data hiding based on erasure and error correction*“*, *IEEE Trans. on Image Processing*, vol. 13, pp. 1627 –1639, Dec. 2004.

11) K. Solanki, A. Sarkar, and B. S. Manjunath, "YASS: Yet Another Steganographic Scheme that Resists Blind Steganalysis", in Proc. *9th International Workshop on Information Hiding, Saint Malo, Brittany, France*, Jun. 2007.

12) K. Solanki, A. Sarkar, and B. S. Manjunath, "Further Study on YASS: Steganography Based on Randomized Embedding to Resist Blind Steganalysis", in Proc. *SPIE - Security, Steganography, and Watermarking of Multimedia Contents (X), San Jose, California*, Jan. 2008.

13) H. Farid, and L. Siwei, ``Detecting Hidden Messages Using Higher-Order Statistics and Support Vector Machines", in Proc. *5th Int. Workshop on Information Hiding, Noordwijkerhout, The Netherlands*, pp. 340-354, 7–-9 Oct. 2002.