J. Basic. Appl. Sci. Res., 3(1)543-549, 2013
© 2013, TextRoad Publication

ISSN 2090-4304
Journal of Basic and Applied Scientific Research
www.textroad.com

Improvement of Steganography Parameters in Image against Steganalysis

Mohsen Jahanshahi[1], Mojtaba Hosseini[2]

[1]Department of Computer Engineering, Central Tehran Branch, Islamic Azad University, Tehran, Iran
[2]Department of Computer Engineering, Amirkabir University of Technology, Tehran, Iran

ABSTRACT

According to the daily increment application of steganography, identification of attack methods to the suspicious media has become to an inevitable affair in order to secret message detection (steganalysis).Different methods of analysis try to represent the difference between natural images and steganography images, by various images of statistical evaluation. Then, based on this difference and usually by using this intelligential methods and training of the appropriate classifier system, gain the model for breakup steganography images from natural images. In this article, the relatively comprehensive system will introduced for nine video formats, namely BMP, JPG, TIF, GIF, PGM, PNG, SWF, and JPEG2000 which has done optimal analysis methods by considering different ways of steganography. The main components of system are the number of SVM classifier that are trained for each video format on a certain number of statistical features. When one secret image arrives into this analyst system, decision making about it will be done based on classifier vote in a hierarchical structure. Experiments and training have been conducted with more than 600 images on the data base and on average more than 85% correct accurate diagnosis have obtained for steganography methods and different embedding rate, and less than 10% for improperly error diagnosis. These values in compared with other methods of rival are so significant, yet the proposed system, due to universalizability is better than other methods.

Keywords: Steganography, Steganalysis, Detectability, Image carrier, Image format

INTRODUCTION

Steganography is the art of embedding one secret message in digital media, like image(especially when it passes from data network) as detection of it, which is not too easy. In recent years many developments have been done about the steganography in many kinds of video formats. In contrast, attack methods to suspicious media (steganalysis) are advancing so rapidly. In the most cases, steganalysis is to detect the presence or absence of secret message in the digital media (image),so it is deemed as a classification problem. In general, the approach which is taken to train the right features for classifications and steganalysis depends on different cases that we can point to two cases:

1. Video format and coefficients domain in steganography:
   Coefficient domain is data which can be saved in a video file and sometimes it is compressed as a video file, and through these can repair the original image. These cases are the location domain coefficients (RGB values or pixel-radiation intensity) TIF, PGM, BMP formats, Discrete Cosine Transformation coefficients (DCT)in JPG and flash(SWF). Discrete violet coefficients in JPEG2000 formats, and also indexes or pixel values in palette images like GIF format and many different palette formats of MNG, PNG, TIF, BMP.
2. Steganography methods:

Various algorithms of steganography in image utilize above coefficients to embedding secret message in different ways. In between, steganography methods with the aim of enhancing the existing analysis methods are introduced every day that compared to the existing analysis methods are more stable so the analyst should have the relatively adequate knowledge about the specific steganography methods and also the style of attack to them.On top of, the accuracy of analysis methods depends on some factors such as being or not being compressed of one carrier image, colored or black and whites, Embedding rate, also it depends on the arrangement of secret messages that can be sequential or random. In addition, one of the important elements in the design of a general analysis system, is the collection of video data that is used to train, and also a collection of images which have a large variety in point of view complexity (text turf), quality, compression ratio, and dimension is required.

To obtain reliable and documented results, images should be embedded with different methods of steganography. To train and test the system, the number of images should be high enough, there are many different protocols and techniques that make us able to hide data into an object data (files). Nevertheless, all of the techniques

***Corresponding Author:** Mohsen Jahanshahi, Department of Computer Engineering, Central Tehran Branch, Islamic Azad University, Tehran, Iran.mjahanshahi@iauctb.ac.ir

and protocols should provide some of requirements, till steganography to be accomplished correctly. A list of main requirements, those steganography techniques should provide, written as below:

- Secret data after embedding into the carrier file should be correct and fault-free. Secret message must not be changed one what so ever, and another information should not be add or subtract.
- Carrier file of message should be remain unchanged (or almost unchanged) toward the naked eye. If a significant change occurs in the carrier file, third party may understand that some information are hidden in this file, and try to extract or destruct them.
- The changes over the carrier file should not have an effect on sign. Suppose you have a legal copy of a picture that you would like to manipulate in different ways, these changes can consist of simple processes such as changing dimensions (size), cutting of redundancies or image rotation. Image sign must be preserved after manipulations, otherwise rivals will be able to delete signs easily and finally steganography will fail.

The final point is here that we always assume attacker is aware of data inside the carrier file. In this regard, a detailed study was conducted about various video formats and also steganography methods including BMP, TIF, PNG, MNG, PGM,GIF, JPG,SWF,JPEG2000.[1,4,9,20,23]. Then various ideas of steganalysis are observed for attack to different algorithms of secret message embedding. Finally optimization analysis methods are designed based on effects of various steganography methods on image, and all generated distortions which embedded in the body of analysis system, and can be used for mentioned video format.

The rest of this paper is organized as follows: in section 2, the reported works have been surveyed. The proposed system is introduced in section 3. Section 4 is dedicated to performance evaluation of the propose scheme as compared with many well known reported approaches. Section 6 concludes the paper and presents some hints for future researches.

2. RELATED WORKS

Steganalysis methods can be generally divided into five groups based on data domain that is used to embed the message and the attack algorithm.

These five groups are used for different video formats as follows:

- General analysis methods (independent of the image format and steganography methods, usable for all formats) [13, 16,19].
- Analysis methods of location domain (related to MNG, TIF, PNG, PGM, BMP)[1,2,5].
- Analysis methods of the DCT domain (related to flash, JPG images)[4,8,2].
- Analysis methods of palette domain (related to GIF images and various palette formats of MNG, PNG, TIF, BMP)[21, 22].
- Analysis methods of coding domain(related to LZW coding process in GIF and TIF formats and also LZ 77 coding in MNG and PGN, TIF formats)[23, 25].

The idea of analysis is used in one of these ways based on what kind of method is used to hide information and how this distortion affects image.

3. THE PROPOSED ANALYST SYSTEM

After a comprehensive perusal on all methods of steganography and steganalysis, a general framework is presented for attacking to steganography in nine formats including BMP, TIF, PNG, MNG, PGM, GIF, JPG, SWF, and JPEG2000.That can detect suspicious steganography image with a high accurate by using of them. Thus the system is organized in such a way that automatically select a suitable approach for analysis based on images format and applied domain at the real time of steganography. This leads to increase accuracy and speed. The main components of analyst system (Fig.1) and their applications are summarized as the following aspects:

- Domain-location analyst(RGB block):

This step is designed to attack steganography at domain location for colored images or black and whites in MNG,PNG, PGM, TIF, BMP formats, and consists of two main parts that follow each other in the form of hierarchical. Module 1 mainly is designed to attack replacement in LSB and module 2 to attack LSB harmony and also other methods of steganography in the place of designed.

- DCT domain analyst (DA):

This step is designed for attack to steganography in DCT coefficients domain in JPG and SWF formats, and consists of two main parts that they follow each other in the form of hierarchical. Module 1 mainly is designed for attack to steganography by the way of harmony LSB and DCT coefficients and module 2 for attack to steganography by the way of replacement in DCT coefficients, and all of the methods which effects on first-order statistics.

- Palette domain analyst:
  This step is designed for attack to steganography in domain location for colored images or black and whites in GIF formats and various palette formats, such as BMP, TIF, PNG, MNG, and consists of two main parts that follow each other hierarchical. Both module of this section, use general methods features based on coefficients of higher-order statistics.
- Coding-domain analyst(CA):
  Already this section has a module designed for attack to steganography during LZW coding. This section is usable for GIF and TIF formats.
- Wavelet coefficients domain(WA)2:
  This step consists of one module for steganography in wavelet coefficients that used in JPEG2000 format and worked based on statistics measuring.

In the block diagram of Fig.1, the general structure of analysis system is shown. Each blocks of BMP, TIF, JPG, etc. are made up smaller components that mainly consist of these five main steps.

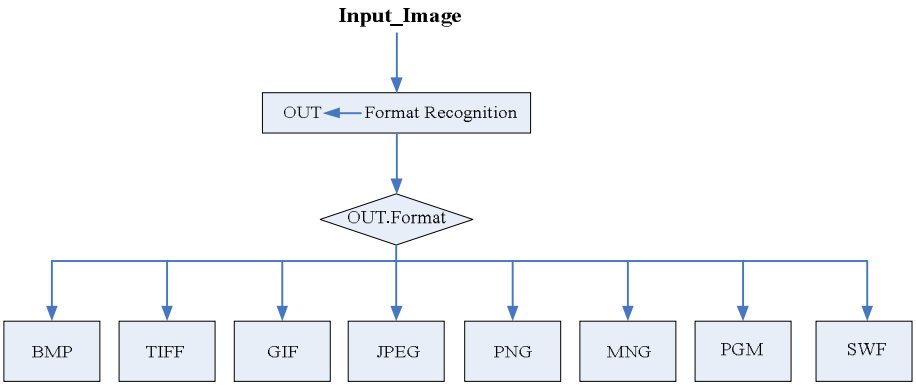


Fig1.The block diagram of the proposed analyst system

The selection of hierarchical structure and use of different features in many parts of system, are organized generally intelligently in order to follow the aim of the whole system. This structure causes both accuracy and speed of decision are increased at the same time because it avoid unnecessary calculations in some of the cases that can recognize some of the steganography methods just only with one group of the features. Hierarchical decision, in each step, is like such a this way that if first module voted to steganography image, trust it, and system announces total vote (based on steganography of input image), otherwise image arrives into the second module and more studies about it will be conducted. In every case, the reason of trusting the first module is the high accuracy of correct diagnosis (TP) and very low rates of error misdiagnosis.

4. EXPERIMENTAL RESULTS

In this section, the proposed scheme is evaluated. For this, a large pictorial data base has been used which its details come in next sub-section.

4.1. Pictorial data bases

These pictorial data bases consist of images with dimensions of about 3000*2000 then they are saved in TIF uncompressed format among the available images, about 2000 shear images are saved in BMP format with dimensions of 720*960 and 960*720 and 670*500 and 500*670.Then these images are converted in JPG format with different quality coefficients(99 and 90, 83, 77,69,52,45,25). In total number of 16000 cover image are produced in JPG format with 8 quality coefficients. And also 2000cover images in BMP format. Among these 2000

images in each collection, 1000 images are considered for training of SVM model and about 1000 images for testing. These images are so various in point of view complexity. About 86 GIF images are used to training of the palette module in which about 500 of them are BMP natural images that are saved in the form of GIF again, and the other 360 images are artificial images that are downloaded from internet.

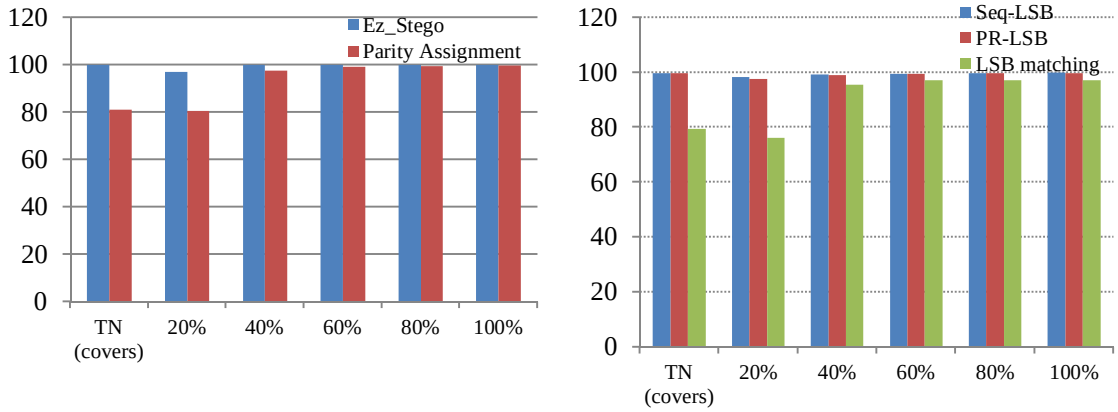
To produce proper samples of steganography images in each format, various steganography methods are implemented.

These methods are summarized by table2.

4.2. Practical results on database images

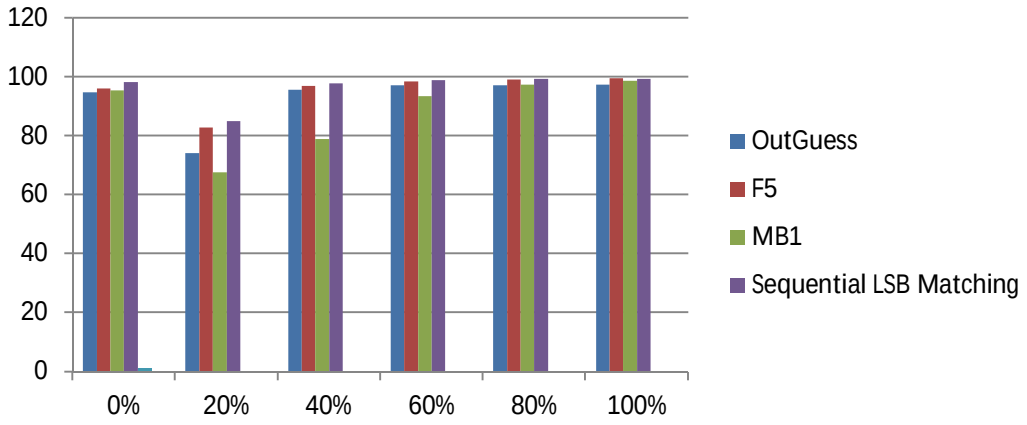
About 1000 cover images and 1000 selective steganography images (randomly) are utilized to train embedded SVM model which are attached to that section of system.

In each case after training of classifier, the power of generalization and accuracy have been estimated by 1000 remaining cover images and corresponding steganography samples (in embedding rate and different methods)test results are shown in the diagrams of fig.1 and fig.2.

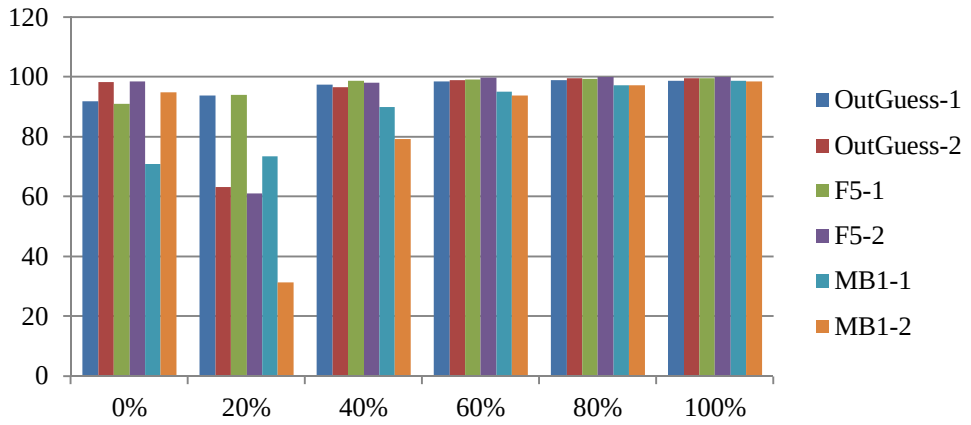


(A) **(B)**

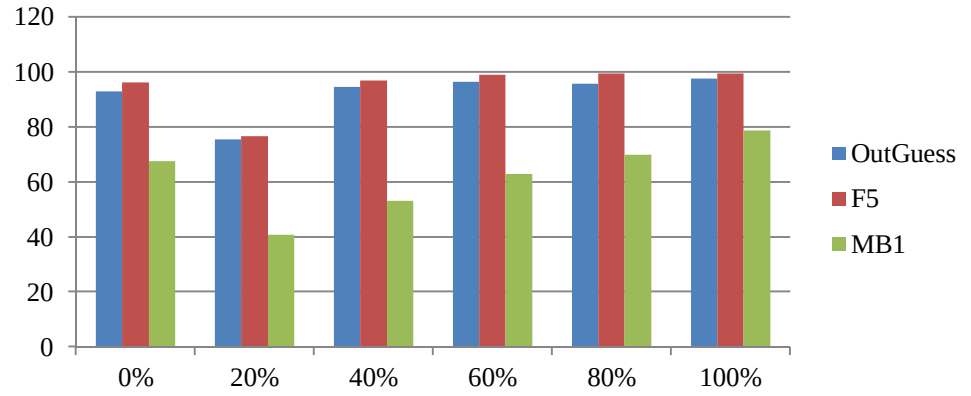
Figure2: Correct diagnosis percentage for test images in different replacement rates.(A). EZ steganography methods compared to parity allocations. (B). Three steganography methods in domain location in LSB and also the average of accuracy to accommodate sequential LSB and random.



A) final results of system for attack to OUTGUESS, F5, MB1 and sequential LSB matching steganography methods in collections of images with quality factor of 77.



B) the results of the reference method[4] to individual training on each method and each embedded rate (state1), and also training on an embedded rate that are attached to one method and test on the same methods(state2).



C) The results of the presented methods in the [3] to training on one specific method in different embedded rate and test on each image.

Fig. 3: diagrams of correct diagnosis for JPG test and comparison of them with result of other researchers in different inserting rates.

Table1: the details of providing pictorial database such as the methods of steganography are used in training of each classifier in different modules of system embedded rate and many cases that used in different formats are listed in this table.

No:	Input format for learning	Used format in learned modules	Steganography methods used in created database	Total image produced in the database	Insertion rate
1	BMP	BMP, TIF, PGM, PNG, MNG	sequential and random LSB Insertion	40000	20% 40% 60% 80% 100%
2	JPG	JPG ،SWF	JSTEG, OUTGUESS, MB1, F5 (sequential and random LSB Insertion in DCT Coefficient)	576000 (8 level in quality)	
3	GIF	GIF, BMP, TIF, PNG, and MNG	EzStego, Parity Assignment	8600	
4	JPEG2000	JPEG2000	LSB Insertion and wavelet coefficient adaptation in sequential and random modes	2000	

5. CONCLUSION

In this article, an optimal method for steganalysis in many various video formats was proposed that outperforms the existing methods. In this approach, the effects of different process of steganography are analyzed, and these

effects and distortions are categorized among the different steganography algorithms. Then, to detect these distortions and create distinction between two groups of natural images and steganography a collection of optimal features is selected that based on a specific logic determine if being or not being steganography image. This system has more superiority than competing system, in selection process and category of appropriate features to steganalysis. Moreover, in the training and testing process, many advantages can be cited for it; First, system universalizability. Second, select proper methods of feature extraction and optimization of these features. Third, choosing hierarchical structure in order to enhance the quality. Although, in this study while achieving appropriate universalizability for various methods of steganography, accuracy is maintained considerably, but in some steganography methods (like MB1) can do more improvement on accuracy by inserting newest modules. The aim of this paper was to introduce a general analyst system, but using a structure design it can be able to identify the methods after initial detection of steganography schemes group, by inserting newest modules to it (which work based on statistical features). This issue can be more important for lower embedding rates.

Acknowledgment

The authors gratefully acknowledge the financial and other support of this research, provided by Central Tehran Branch, Islamic Azad University, Tehran, Iran.

REFERENCES

[1] J. Fridrich, "Feature-Based Steganalysis for JPEG Images and its Implications for Future Design of Steganographic Schemes", *Information Hiding, 6th International Workshop, LNCS 3200,* PP 67-81, 2004.

[2] Tomas Pevny, Jessica Fridrich, "Multi-Class Detector of Current Steganographic Methods for JPEG Format", IEEE, 2008.

[3] Tomas Pevny, Jessica Fridrich, "Merging Markov and DCT Features for Multi-Class JPEG Steganalysis", SPIE, 2006.

[4] Dongdong Fu, Yun Q. Shi, DekunZou, GuorongXuan, “JPEG Steganalysis Using Empirical Transition Matrix in Block DCT Domain”, Department of Electrical and Computer Engineering, New Jersey Institute of Technology, 2005.

[5] Xiang-Yang Luo, Dao-Shun Wang, Ping Wang, Fen-Lin Liu, “A review on blind detection for image steganography”, Signal Processing, doi:10.1016/j.sigpro.2008.03.016, (www.elsevier.com/locate/sigpro), 2008.

[6] S. Lyu and H. Farid, "Detecting Hidden Messages Using Higher-Order Statistics and Support Vector Machines" Proc. 5th Inrernarional Workshop on Informarion Hid- ing, 2002.

[7] TarasHolotyak, Jessica Fridrich, SviatoslavVoloshynovskiy, "Blind Statistical Steganalysis of Additive Steganography Using Wavelet Higher Order Statistics", CMS2005.

[8] J. Fridrich, M. Goljan, D. Hogea, "Steganalysis of JPEG Images: Breaking the F5 algorithm", 1 Department of Electrical and Computer Engineering, SUNY Binghamton, Binghamton, NY 13902-6000, USA, 2003.

[9] http://sprott.physics.wisc.edu/fractals.htm

[10] Jeremiah Joseph Harmsen, “STEGANALYSIS OF ADDITIVE NOISE MODELABLE INFORMATION HIDING”, MASTER OF SCIENCE Thesis, Rensselaer Polytechnic Institute Troy, New York, April 2003.

[11] Qingzhong Liu, Andrew H. Sung, BernardeteRibeiro, Mingzhen Wei, Zhongxue Chen, JianyunXu, "Image complexity and feature mining for steganalysis of least significant bit matching steganography", Information Sciences, No. 178, pp. 21–36, 2008.

[12] Andreas Westfeld and Andreas Pfitzmann, "Attacks on Steganographic Systems, Breaking the Steganographic Utilities EzStego, Jsteg, Steganos, and S-Tools—and Some Lessons Learned", Dresden University of Technology, 1998.

[13] A. Ker, “Steganalysis of LSB matching in grayscale images,” IEEE Signal Process. Lett., vol. 12, no. 6, pp. 441–444, June 2005

[14] M.U.Celik, G.Sharma, A.M. Tekalp, "Universal Image Steganalysis Using Rate-Distortion Curves", Proc SPIE : Security, Steganography and Watermarking of Multimedia Contents VI, vol. 5306, San Jose, 19-22, Jan, 2004.

[15] Hongmei Gou, AshwinSwaminathan and Min Wu, “NOISE FEATURES FOR IMAGE TAMPERING DETECTION AND STEGANALYSIS” , ECE Department, University of Maryland, ICIP, 2007.

[16] SorinaDumitrescu, Xiaolin Wu, Zhe Wang, "Detection of LSB Steganography via Sample Pair Analysis",IEEE TRANSACTIONS ON SIGNAL PROCESSING, VOL. 51, NO. 7, pp. 1995-2007, JULY 2003

[17] J. Fridrich, M. Goljan, R. Du, "Detecting LSB Steganography in Color and Gray-Scale Images", Multimedia and Security Magazine, IEEE, Oct-Dec 2001

[18] S.Mitra, T.Roy, D.Mazumdar, A.B.Saha, "STEGANALYSIS OF LSB ENCODING IN UNCOMPRESSED IMAGES BY CLOSED COLOUR PAIR ANALYSIS", cp03

[19] Andrew D. Ker, "A Weighted Stego Image Detector for Sequential LSB Replacement", Oxford University Computing Laboratory, 2007

[20] Jiaohua Qin, Xingming Sun, Xuyu Xiang, "Steganalysis Based on Lifting Wavelet Transform for Palette Images", 2007 International Conference on Computational Intelligence and Security Workshops, pp. 672-675. 2007.

[21] Jessica Fridrich, MiroslavGoljan, David Soukal, "Higher-order statistical steganalysis of palette images" , in Proc. SPIE, Security andWatermarking of Multimedia Contents V, E. J. Delp III and P. W. Wong, Eds., 2003, vol. 5020, pp. 178–190.

[22] Xinpeng Zhang, Shuozhong Wang, "Analysis of Parity Assignment Steganography in Palette Images", Springer-Verlag Berlin Heidelberg 2005, KES 2005, LNAI 3683, pp. 1025–1031, 2005.

[23] Hiuk Jae Shim, JinhaengAhn, ByeungwooJeon, "DH-LZW_ LOSSLESS DATA HIDING IN LZW COMPRESSION", 2004 International Conference on Image Processing (ICIP).

[24] Xiaochuan Chen, Yunhong Wang, Tieniu Tan, Lei Guo, "Blind Image Steganalysis Based on Statistical Analysis of Empirical Matrix", The 18th International Conference on Pattern Recognition (ICPR'06), IEEE 2006.

[25] HediehSajedi and Mansour Jamzad, "A Steganalysis Method Based on Contourlet Transform Coefficients", International Conference on Intelligent Information Hiding and Multimedia Signal Processing, IEEE, 2008, pp. 245-248.