ISSN No. 0976-5697

Volume 8, No. 5, May – June 2017

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Trapping of Stego Images on the Basis of Statistical Evidences

Pammi Anusha
Department of Information Technology
Bharati Vidyapeeth Deemed University,
College of Engineering
Pune, India

Prof.S.P. Medhane
Department of Information Technology
Bharati Vidyapeeth Deemed University,
College of Engineering
Pune, India

Abstract: Image-based steganography is an information hiding technique for improving information security. Its purpose is to cover the original data in image to secure the data. Recently, a stego-image detection scheme was proposed, which uses the so-called extreme learning machine and features of digital images for analysis, hiding data through steganography and analysing it through steganalysis which is used to find whether an image contains secrete data. Steganography is developed to hide the data using some digital media and steganalysis is explained as the technique to find the hidden data in the digital media; can also be explained as analysing the steganography presence. The information can be hidden in images in different domains like discrete cosine transform and spatial. The algorithm changes the properties of the image due to embedded artefacts. In this project the main goal is to develop a steganalysis system to identify the presence of hidden information in images, based on Image Quality Measures as well as identify the steganography embedding domain using Support Vector Machine.

Keywords: Steganography, Steganalysis, Support Vector Machine, Classifier.

I. INTRODUCTION

Steganography is explained as the technique to hide the data in such a way that it cannot be detected by human eye. The sender sends the data by embedding the data bits in digital media in a way that no one except the sender knows that data is hidden in that digital media [1]. Steganography means "covered writing" in Greek. The aim of steganalysis is to analyze digital media to check the presence of hidden data that cannot be detected by human eye.

Steganalysis can be used to break the security of the data that is hidden in digital media. For steganographic methods all formats of digital files are suitable and most probably digital images and audio files are used as they provide redundancy of high degree.

There are number of techniques used to hide data but image steganography is the popular one. In this method of steganography, the data that is to be sent secretly is converted into bits and embedded in the bits of an image called cover image and a human eye cannot find the data or detect the changes in the image. To embed the data bits into image bits there are many techniques proposed to hide the data securely. There also exist techniques at receiver side to extract the main message data from the image bits. If audio is cover media, then the data is embedded as noise in audio file that cannot be heard by human ear as its range is high. There exists another method of steganography which uses language that is used naturally to find the secret data, and is called text steganography or linguistic steganography [2]. The process of analysis of cover media which results in separation of secret information from cover media is called as steganalysis [3].Various steganalysis techniques are proposed in literature for detection of hidden information in cover media like image, text, audio and video. Video steganography has capabilities to carry maximum secret data than in other technique like image steganography or audio steganography. If embedding approach is known for steganalysis then it is called as Targeted Steganalysis. If embedding approach is not known, then it is termed as Blind Steganalysis.

There are many techniques and domains used to hide data in to the digital media and some of them are as follows. One of most frequently used technique is substitution process in which the data bits that are to be sent secretly are substituted in least significant bit position of a pixel. This process also contains many drawbacks in keeping the data secretly. The second method that is popularly used in steganography is transform technique which uses various methods of transformations such as discrete cosine transform [4], discrete wavelet transform and the third transform technique used is Fast Fourier Transform. These transform techniques hide the data in coefficients called transform coefficients of the digital media. Spread Spectrum Technique and Statistical Technique are two most frequently used techniques. In Spread Spectrum Technique bandwidth of frequency of a range is used for sending data secretly. Statistical Technique uses blocks of fixed size of the image that is used to cover the data that is to be sent secretly. There are techniques used in steganography that uses image [5] as its media. These are Pixel differences based measures, correlation based measures, spectral distance based measures, edge based measures, and context based measures.

Steganalysis is tool reinforced steganography. Steganalysis is classified into two main categories [6]. The first type of steganalysis is signature method and the second approach is statistical analysis [7]. This type of classification depends on signature of the steganography and the image statistics in which data bits are embedded in the media [8]. Statistical analysis, as the name itself explains, checks the original image statistics to analyze the information embedded in to the image [9]. Steganalysis of statistical type is measured influential compared to that of signature analysis since precise methods are known to be more delicate that of graphical [10-11]. Steganography method coats secretive information by working on digital media such as images, audio, and video to keep data invisible to naked eye. Hiding data in any of cover media with the help of steganography method needs a change to the properties of cover media. The existence of patterns and characters are used as signatures to analyze the presence of message [12, 13]. Steganalytic approaches can be separated into targeted and blind steganalysis. Predictor must develop a

steganalysis algorithm that is proficient of sensing all forms of steganography. This type of steganalysis is known as blind steganalysis. In other case, if predictor able to detect a definite steganography algorithm; it is termed as targeted steganalysis. The hierarchy of the classification of steganography techniques is explained in detail [14].

II. LITERATURE REVIEW

There are many methods proposed to implement steganography. The main goal of each method is to secure data and send to the receiver in such a way that no other person can understand that the digital media contains data.

A. Steganographic Methods

Some of the methods are discussed below:

1) Data Hiding by LSB: Many methods for information hiding have been anticipated in literature. Most frequently used technique to embed the data into image bits is to insert the data bits in to the least significant bits of the pixel of an image. This insertion makes the slight changes in the image which are not recognised by naked eye.

2) JSteg: This method is a technique used in previous days. As per this method the image of JPEG is used as the media to embed the data bits into the least significant bits. This process hides the data in image called cover image altering the least significant bits of discrete cosine transform coefficient which has non-zero quantized index with data bits.

3) F5: In this method, the message bits are inserted by considering discrete cosine transform coefficient that is chosen randomly. This method uses bits' matrix for embedding that is used to reduce the changes required to hide the message.

B. Steganalysis Methods

Steganalysis is done to analyze the presence of secret message in the media. There are many methods proposed from past decade to check for message that is inserted in to image bits.

1) Steganalysis by SPAM: In this method stego signal of low amplitude is taken by taking the difference between pixels that are adjacent and uses first order and second order Markov chain. In this method transition probability matrix is formed which plays an important role for SVM based steganalysis. The transition probability and difference between pixels are determined using eight directions.

2) Steganalysis of YASS: As per this method JPEG steganography images are used for analysis [11]. In this technique, the image is divided into blocks of fixed size. Data is recovered from the image bits after the completion of decompression of image. In this method, Discrete Cosine Transform is performed by using a secrete key that is chosen randomly. It selects domain of steganalysis through feature extraction from image.

3) Digital media Steganalysis by Ensemble Classifier: This method is executed as random forest. It supports multiclass classifier that is trained with different examples by considering various dimensionality features. The training of SVM will be slow as it supports multiclass. Statistical based feature algorithm is used to train the classifier with high order features and calculated from image transformations by filters [14]. Also, considering similar metrics of 18 binary values [15], and features by discrete cosine transform [16]. This method also takes approximately 27 wavelet coefficients of high order [17]. Performance measure must be accurate for steganalysis; and for high accuracy rate, steganalysis should have feature vector which will be combination of all together for high dimension. The calculation of JPEG feature set used in paper cited [18] are 274 and this features are extended by calibration of Cartesian [19]. Total detection error used in ensemble classifier for training set will be as follows:

$$PE= 1/ 2 \min [P_{FA}+ P_{MD}(P_{FA})] \quad (1)$$

III. PROPOSED METHODOLOGY

The main goal of this work is to develop a system which can classify the cover multimedia and hidden data. Objective of this work involves developing a classifier to know whether the image is original or stego image.

In our proposed method, an approach is developed by which trapping of stego image is done and tries to find out the presence of secret message in image. This approach is divided into three parts which will trap the images for analysis.

In first part of proposed work, a standard database which is used for steganography purpose at international level i.e. BOSSBase database and which has more than 10000 digital images whose features are extracted and stored in data file for further processing. This process is called as feature extraction and involves calculating more than 600 features of an image through subtractive pixel matrix.

Second part consists of classifier training through BOSSBase database features that are extracted in first part. Support vector machine is used for classification. It is a binary classifier used for discrimination purpose. In the last part of proposed work, a random sample from database, whose features are calculated by the same algorithm that was used for feature extraction of database images is compared with the standard database through the support vector machine. The result show whether the image has any hidden data in it or not. Using this proposed work, we can detect the presence of hidden data in provided images as well as what kind of stenography technique is used to hide that data.

IV. EXPERIMENTAL RESULTS

Experiment is carried out on BOSSBase database, which has 512x512 images with 75 to 90 % payload. Ensemble classifier is used for discrimination propose.

Ensemble classification:

```
    # - Training samples: 1994 (997/997)
    # - Feature-space dimensionality: 548
    # - L: 30
    # - d_sub: 300
    # - Seed 1 (subspaces): 5
    # - Seed 2 (bootstrap): 73
    # -------------------------
     - d_sub 300: OOB 0.1148: L 30: T 2.0 sec
    # -------------------------
    Optimal d_sub 300: OOB 0.1148: L 30: T 2.0 sec
    OOB error = 0.1053
Average testing error over 10 splits: 0.1021 (+/- 0.0052)
```

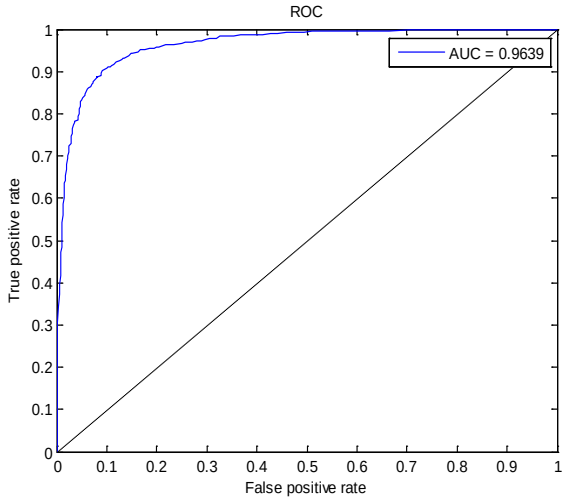

Figure 1. False positive rate

Figure 1 shows the ROC of classifier with accuracy 0.9 and above. The true positive alarm means classifying the stego image as stego and false positive alarm means classifying cover image as cover. Our proposed work gives the accuracy more than 0.95 which is greater than the accuracy of any present technique.

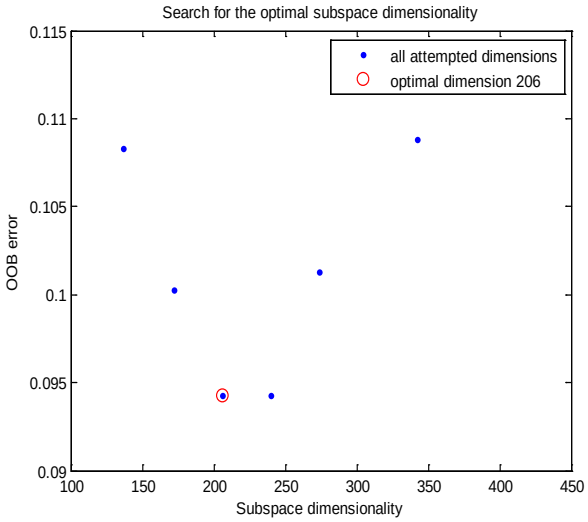

Figure 2. Subspace dimensionality

Figure 2 shows trapping the stego images using the properties of stego and cover images that are determined during computation. Using SPAM calculation of 658 features of an image that can be used for classification by SVM. The main challenge is to optimize the feature set without changing the accuracy of the proposed work. Figure show that we have optimized feature set to 214 with the same accuracy.

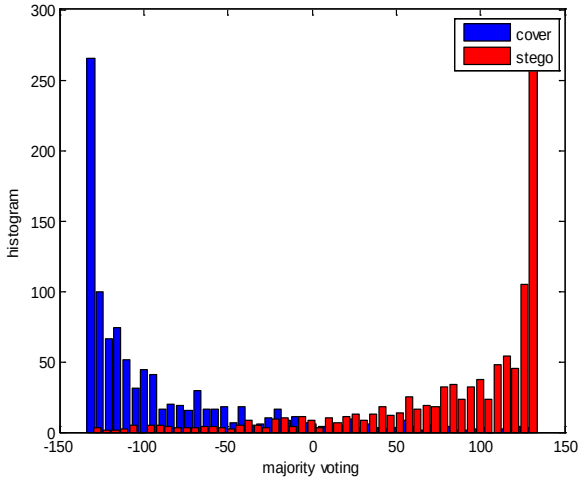

Figure 3. Majority voting

Figure 3 shows cover and stego images for processing. The figure shows the variation in histogram of cover images after embedding secret data in it. Most of the steganography techniques use high entropy area to hide the secret data in images so the variation in cover and stego images histogram is reflected in figure.

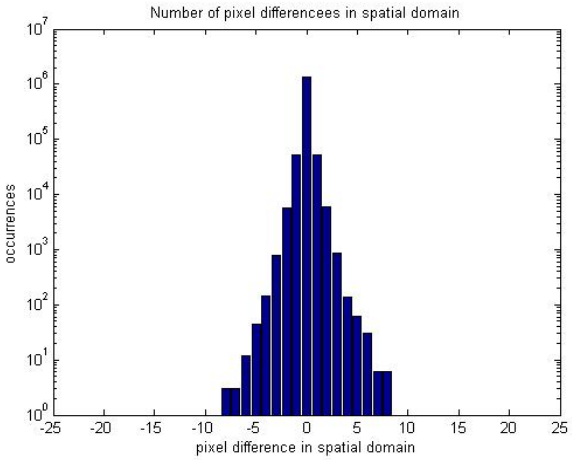

Figure 4. Pixel difference in spatial domain

In figure 4 the image consists of pixels with different density. As shown in fig the high-density pixel will be used to hide the secret data as they are having highest probability to get modified.

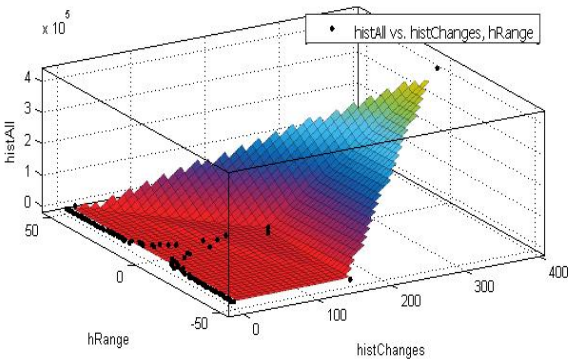

Figure 5. Histogram variation

Figure 5 shows the histogram variation with respect to its pixel density.

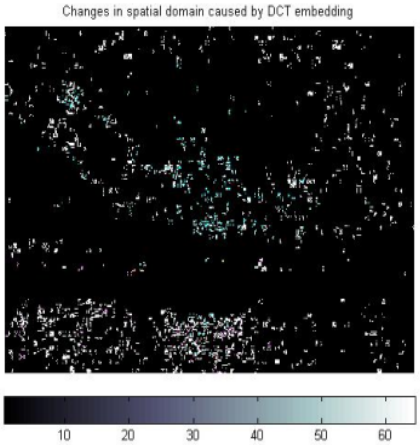

Figure 6. Changes in spatial domain caused by DCT embedding

As seen in Figure 6, the white spots reflect the effect of DCT domain steganography in spatial domain. In DCT domain steganography the secret data is stored into frequency coefficients that can be trapped using spatial domain analysis as shown in Figure 6.

V. CONCLUSION

The proposed work will give new approach for detection of hidden data in digital image. It will demonstrate the ability of classifying stego and cover images. Proposed work will be useful for forensic analysis of digital images and filtering for effective steganalysis.

VI. REFERENCES

[1] G. R. Suryawanshi, Dr. S N Mali, "Study of Effects of DCT Domian Steganography Techniques in Spatial Doamain for JPEG Images Steganalysis", Internationl Journal of Computer Applciaiton, Vol. 127, pp.16-20, October 2015

[2] M.Chapman, G. Davida, and M. Rennhard, "A Practical and Effective Approach to Large-Scale Automated Linguistic Steganography", Proceedings of the Information Security Conference, 2001, pp.156-165.

[3] Udit Budhia, Deepa Kundur and Takis Zourntos,"Digital Video Steganalysis Exploiting Statistical Visibility in the Temporal Domain", IEEE Transactions On Information Forensics And Security, vol.1, pp. 502-516, 2006.

[4] Jessica Fridrich, Member, IEEE and Jan Kodovský,"Rich Models For Steganalysis Of Digital Images", IEEE Transactions On Information Forensics And Security,vol.7, pp. 868-882, 2012..

[5] Min Wu, Member, IEEE, and Bede Liu, Fellow IEEE, "Data Hiding in Image and Video:Part I—Fundamental Issues and Solutions", IEEE Transactions on Image Processing, vol. 12, pp. 685-695, 2003.

[6] Chunfang Yang, Fenlin Liu, Xiangyang Luo, and Bin Liu,"Steganalysis Frameworks of Embedding in Multiple Least-Significant Bits", IEEE Transactions on Information Forensics and Security, vol.3, pp. 662-672, 2012.

[7] Yun Cao, Xianfeng Zhao and Dengguo Feng,"Video Steganalysis Exploiting Motion Vector Reversion Based Features", IEEE Signal Processing Letters,vol.19, pp.35-38, 2012.

[8] Mengyu Qiao, Andrew H. Sung , Qingzhong Liu," MP3 audio steganalysis", Information Sciences 231 Elsevier, pp. 123–134, 2013.

[9] R. Sridevi, A. Damodaram and S.V.L. Narasimham," Efficient Method of Audio Steganography by Modified LSB Algorithm and Strong Encryption Key with Enhanced Security",Journal of Theoretical and Applied Information Technology,vol.5, pp.768 – 771, 2009.

[10] H. D. Kirovski and H. Malvar,"Spread spectrum Watermarking of Audio Signals", IEEE Transactions on Signal Processing,vol.51, pp.1020 – 1033, 2003 .

[11] Tomáš Pevný, Jessica Fridrich and Andrew D. Ker, "From Blind To Quantitative Steganalysis",IEEE Transactions On Information Forensics And Security,vol.7, pp.445-454, 2012.

[12] Jan Kodovský, Jessica Fridrich,"Ensemble Classifiers For Steganalysis Of Digital Media",IEEE Transactions On Information Forensics And Security,vol.7, pp.432-444, 2012.

[13] Jan Kodovský And Jessica Fridrich,"Quantitative Structural Steganalysis Of Jsteg",IEEE Transactions On Information Forensics And Security,vol. 5, pp. 681- 693, 2010.

[14] Abbas Chaddad, Joan Condell, Kevin Curran, Paul Mc Kevitt," Digital Image Steganography",Survey and Analysis of current Methods, ELSEVIER, Signal Processing 90, 2010, pp.727-752.

[15] Yun Cao, Xianfeng Zhao, and Dengguo Feng, "Video Steganalysis Exploiting Motion Vector Reversion Based Features", IEEE Signal Processing Letters,vol.19, pp.35-38, 2012.

[16] Tomáˇs Pevný, Patrick Bas and Jessica Fridrich,"Steganalysis By Subtractive Pixel Adjacency Matrix",IEEE Transactions On Information Forensics And Security,vol.5, pp. 215-224, 2010.

[17] Yu Deng, Yunjie Wu, Haibin Duan, Linna Zhou,"Digital Video Steganalysis based on Motion Vector Statistical Characteristics",ELSEVIER,Optik, pp. 1705-1710,2012.

[18] Matthew C. Stamm, W. Sabrina Lin, And K. J. Ray Liu,"Temporal Forensics And Anti-Forensics For Motion Compensated Video",IEEE Transactions On Information Forensics And Security,vol.7, pp. 1315-1329, 2012.

[19] Yong Wang, Jiufen Liu, Weiming Zhang, Shiguo Lian,"Reliable JPGE Steganalysis based on Multi-directional Correlations",ELSEVIER,Signal Processing : Image Communication 25, pp.577-587,2010.