Development of BlindSteganalysis using Co-occurrence Features

SwagotaBera
Department of E&Tc
SSCET, Bhilai
swagotaberasarkar@gmail.

Dr. Monisha Sharma
Department of E&Tc
SSCET, Bhilai
monisha.sharma10@gmail.com

Dr.Bikesh Singh
Department of Bio-Medical
Engineering NIT, Raipur
bsingh.bme@nitrr.ac.in

Abstract

Steganalysisis the technique for the detection of hidden information in an image. Blind Steganalysisis a versatile detection technique whichcanbeimplemented for any type of hiding.Due to hiding the statisticalfeatures of the corner pixel of image getdeviates. Thesefeaturescapture the variations in the image pixels due to hiding. In thispaper, the blindhiding technique wasproposed in which the deviations in the co-occurrencestatisticalfeatures of the corner pixelsof an image wascalculated in DiscreteWavelet Domain (DWT) domain. The proposed technique wasimplemented on the wellknown JPEG steganographic techniques i.e. JUNIWARD, nsF5, WT Based, Outguess, F5 and Jsteg. The Support Vector Machine(SVM) and J48 are the classifiers deployed for the datamining of the obtained features. The sensitivity, specificity and detection accuracy of the proposed technique was found out.

Keywords—Steganography, Steganalysis,DWT, SVM

I. Introduction

Steganalysis is the reverse technique of steganography. The hidden data whether it is a text or image and in any format can be detected.As per the history of this steganography technique, it is found that the terrorist are using this hiding technique for communicating the secret information since last 15 years. Since passed years, various techniques were implemented to stop the secret data transmission by developing and implementing various hidden data detection tools. Though there is vast development in the designing of the hiding technique, in the same space the improvement in the detection technique is also going on.For the favour of society and country and for the whole world, the work in the hidden data detection technique will be beneficial.

Jsteg [3], outguess[1], F5[2], WT Based[4], nsF5[16] and JUNIWARD [17] are the hiding techniques implemented in the present work for generating the stego image database. The image which contains the secret data is known as stego image. Jsteg[3] is JPEG hiding technique in which the zero and one coefficient is not used for hiding. Outguess[1], embeds hidden information into the redundant bits of data sources. It preserves the global histogram of BDCT by adjusting untouched coefficient such that histogram does not change after data hiding. F5[2] modifies the block-DCT coefficients to embed messages. This technique is based on straddling and matrix coding. Straddling scatter the message as uniform distribution and matrix coding improves embedding efficiency. WT based steganography hides the secret data bits in the wavelet coefficients such that the global histogram is preserved after hiding [4]. nsF5[16] and JUNIWARD [17]are content adaptiveembedding technique which is independent of the embedding domain.

Fridrich has proposed a set of distinguishing features from the BDCT domain and spatial domain from the calibrated image [8]. In [10], Shi et.al.presentsuniversal steganalysis by calculating the statistical moments of characteristic functions of the image from its prediction-error image in DWT domain. This steganalyzer provides a better performance.In[6,12,13], the inter-pixel and intra-pixel dependencies are used as features by Markov chain model. In [12], Mahindra Kumar improved the blind steganalysis technique using the two step transition probability matrix and extended DCT features in DCT domain. In [14], co-occurrence matrix and statistical moment is calculated from the contourlet coefficients of the image and further SVM is used for classification.

The work discussed here is based on the designing of a blind detection scheme for the gray scale JPEG image in transform domain.If we analyze any image in detail, it is the finding that the corner pixel values of the objects within an image also carry the statistical information about the image.The statistics of the image get changed due to hiding the secret data. The DWT is the domain transformation mathematical tool which represents the image in both spatial and frequency domain.The wavelets obtained after DWT transformation carries the image information. The coefficient dimension depends on the various levels of orientations and scales that can be selected according to the requirement.The performance computation of the detection scheme is done by using SVM and J48 classifiers.

The organization of the paper is as follows. The image feature generation technique is discussed withmathematical explanations in section II. The overview of the proposed technique is discussed in section III. In section IV experimental setup is presented and in section V, experimental results and conclusions are placed.

II. Image Feature Generation

A. Corner Characteristics of an Image

The large variations in the neighbourhood pixel values are captured in the corner matrix. The corner detection block can be findout in an image using the Harris corner detection, which maintains the trade-off between accuracy and computational complexity.The corner matrix dimension is similar to the image dimensions for the grayscale image.

Let A, B and C is the parameters calculated using the equation (1).

$$A = (I_x)^2 \otimes w$$

$$B = (I_y)^2 \otimes w$$

$$C = (I_z)^2 \otimes w \qquad (1)$$

Where I_x and I_y are the gradients of the input image I in the x and y direction, respectively. The $\otimes$ symbol denotes a convolution operation.Harris corner detection calculates the corner matrix using the formulae in equation (2).

$$R = AB - C^2 - k(A+B)^2 \qquad (2)$$

The variable k corresponds to the sensitivityfactor. You can specify its value using the sensitivity factor (0< k <0.25) parameter. The smaller the valueof k, the more likely it is that the algorithm can detect sharp corners. The default value is 0.04.

B. Co-occurrence Matrix

Co-occurrence matrix is defined over an image as thedistribution of co-occurring values at a given offset. Mathematically, a co-occurrence matrix c is defined overan $m \times n$ image I, parameterized by an offset $(\Delta x, \Delta y)$ as in equation (3).

$$c\Delta x \Delta y(x, y) = \sum_{p=1}^{n} \sum_{q=1}^{m} X$$

Where $x = 1$; if $I(p,q) = i$ and $I(p + \Delta x, q + \Delta y) = j$

$x = 0$; Otherwise (3)

The co-occurrence matrix is determined using a set of 4 offsets sweeping through 180 degrees (i.e. 0, 45, 90, and 135 degrees) at the same distance to achieve a degree of rotational invariance. Also four statistical features i.e. homogeneity, contrast, Angular Second Moment (ASM), and correlation are extracted from co-occurrence matrix of DWT coefficients of corner matrix. Homogeneity measures the closeness of the distribution of elements in the matrix. Contrast shows the amount of local variation present in an image. ASM is a measure of uniformity of theimage and correlation is a measure of gray-tone linear dependencies in the image.

$$Homogeneity = \sum_{i,j}\left(\frac{1}{1+|i-j|}\right)p(i,j)$$

$$Contrast = \sum_{i,j}|i-j|^2\ p(i,j)$$

$$Energy(ASM) = \sum_{i,j} p(i,j)^2$$

$$Correlation = \frac{\sum_{i,j}(i,j)p(i,j)-\mu_x\mu_y}{\sigma_x\sigma_y} \qquad (4)$$

Where σ_x and σ_y are the standard deviations and μ_x and μ_y are the means of p_x and p_y respectively. $p(i,j)$ is the (i,j) the entry in a normalized gray tone spatial matrix and $p(i,j)=\frac{p_{\Delta_x\Delta_y}(i,j)}{R}$, where R is the total number of pixel pairs (i,j).The gray-level co-occurrence matrix is calculated. The value (i,j) in the matrix denotes that how many times the pixel with value i occurs horizontally adjacent to a pixel of value j [14].

III. Methodology

In the proposed technique, the co-occurrence of isused as statistical features extracted from the corner matrix. The DWT is used as transformation function to the image. Finally the classification and performance analysis is performed using classifiers. The proposed scheme is represented in figure(1).

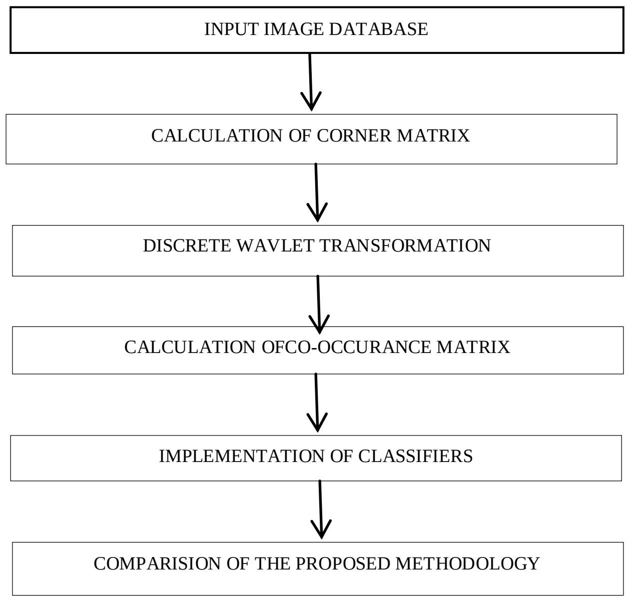

FIGURE (1) BLOCK DIAGRAM OF THE PROPOSED SCHEME

IV. Experimental Setup

A. Image Set

The image set of2500 JPEG images of quality factor 80 were used for generating stego and non-stego images. The images are obtained from Greenspun image library and Google websites. All the images are of resolution 640 x 480.

B. StegoImages Generation

The images from the database are gone through various well known JPEG hiding techniques. They are JUNIWARD, nsF5, WT Based, Outguess, F5 and Jstegof different capacities 0.05, 0.1, 0.2. The texts and images are hidden in the image dataset using the above algorithm [1,4,18].

C. Steganalysison the Stego Image Database

Matlab code is generated for implementing the feature extraction. The obtained features are classified with the help of WEKA data mining software [11]. The cross-validation is selected for the better result.The performance parameters of classification are shown in the Table I.

III. Results and Conclusions

The detection accuracy is more in case of Jsteg and sF5 as compare to others. JUNIWARD is more advanced hiding technique and hence the detection accuracy is low comparatively. Regarding the execution time it can be concluded that the time requirement is quite less for J48 classifier as compared to SVM and also detection accuracy is also better in case of J48. It is the finding that the proposed technique gives a good detection accuracy for almost all the JPEG hiding techniques but needs to be changes in the domain for to detect more accurately for the content adaptive JPEG hiding technique.

Acknowledgment

Appreciationgoes to Dr.AyushSinghalPostdoctoral ResearchFellow, NCBI, National Institute of Health, Maryland, PhD, M.Sc. (University of Minnesota, Twin Cities, MN, USA) andB.Tech (IIT Roorkee, India)for giving me the information about Weka Data MiningTool.

TABLE I

PERFORMANCE OF RANDOM FOREST CLASSIFIERS FOR THE PROPOSED TECHNIQUE

CLASSIFIERS		**Support Vector Machine (SVM)**		**J48**	
HIDING METHOD	**EMBEDDING CAPACITY**	**Detection Accuracy (%)**	**Execution Time (Seconds)**	**Detection Accuracy (%)**	**Execution Time (Seconds)**
JUNIWARD [17]	**0.05**	50.72	2.23	50	0.03
	0.1	51.12	3.22	50.12	0.03
	0.2	52.32	3.33	52.6	0.01
nsF5[16]	**0.05**	70.02	5.13	70.72	0.05
	0.1	70.8768	4.99	70.84	0.06
	0.2	70.924	5.16	72.04	0.06
DWT Based[4]	**0.05**	60.236	3.08	60.02	0.1
	0.1	60.6	5.03	60.8	0.1
	0.2	62.84	4.06	61.44	0.1
Outguess[1]	**0.05**	60.6	4.05	60.36	0.21
	0.1	60.88	5.06	60.86	0.1
	0.2	61.76	6.05	61.36	0.04

F5[2]	**0.05**	60.1594	5.06	61.04	0.1
	0.1	60.5523	4.06	61.362	0.1
	0.2	60.8611	3.06	63.52	0.1
Jsteg[3]	**0.05**	60.552	3.06	60.52	0.02
	0.1	60.92	4.05	60.92	0.3
	0.2	61.0338	4.05	61.02	0.2

References

[1] "Attacking the Outguess." [Online]. Available:http://dde.binghamton.edu/publications/acm_outguess.pdf. [Accessed: 14-Feb-2016].

[2] Westfeld A. F5—A Steganographic Algorithm: High Capacity Despite BetterSteganalysis. Proceeding of Fourth International Workshop on Information Hiding.USA, 2001, pp.89-302.

[3] T. Qiao, F. Retraint, R. Cogranne, and C. Zitzmann, "Steganalysis of JSteg algorithm using hypothesis testing theory," in *EURASIP J. Inf. Secur*, vol. 2015, no. 1,Mar 2015, p. 2.

[4] D. Umashankar;, S. Monisha, and S.Bera, "Development and analysis of stego image using discrete wavelet transform," in *Int. J. Sci. Res.*,2013, pp. 142–148.

[5] S. Lyu, Farid H, "Detecting Hidden Messages using Higher-Order Statistics andSupport Vector Machines" in*International Workshop on InformationHiding*, Vol.2578, 2003, pp.340–354.

[6] S. yun Q., C. Chen, and Chen Wen, "A Markow process based approach for effectively attacking JPEG stegnography," in *Information Hiding*, vol. 4437, J. L. Camenisch, C. S. Collberg, N. F. Johnson, and P. Sallee, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 249–264.

[7] G. Kumar, R.Jithin, and D. D. Shankar, "Feature Based Steganalysis Using Wavelet Decomposition and Magnitude Statistics," in *International Conference on Advances in Computer Engineering*, 2010, pp. 298–300.

[8] J. Fridrich,"Feature-Based Steganalysis for JPEG Images and its Implication forFuture Design of Steganographic Scheme", in *International Workshop onInformation Hiding*, 2005, pp.67–81.

[9] S. Lyu and H. Farid, "Steganalysis Using Higher-Order Image Statistics," in *IEEE Trans. Inf. Forensics Secur*, vol. 1, no. 1,2006, pp. 111–119.

[10] C. Chen, Y. Shi, W. Chen, and G. Xuan, "Statistical Moments Based Universal Steganalysis using JPEG 2-D Array and 2-D Characteristic Function," in *International Conference on Image Processing*, 2006, pp. 105–108.

[11] “Machine Learning Project at the University of Waikato in New Zealand.” [Online]. Available: http://www.cs.waikato.ac.nz/ml/index.html. [Accessed: 14-Feb-2016].

[12] M. Kumar, “Steganography and Steganalysis of Joint Picture Expert Group (JPEG) Images”, *Ph.D. Thesis, Universityof Florida. 2012.*

[13] D. Fu, Y. Q. Shi, D. Zou and G. Xuan, “JPEG Steganalysis Using Empirical Transition Matrix in Block DCT Domain”, in *IEEE Workshop on multimedia Signal Processing*, 3-6 Oct, 2006, pp. 310-313.

[14] M. Sheikhan, S. Moin and M. Pezhmanpour “Blind Image Steganalysis via Joint Co-occurrence Matrix and StatisticalMomentsofContourlet Transform “, in10th *International Conference on Intelligent System Design and Applications*,29 Nov, 2010, pp.368-372.

[15] S. Jinyang,Z. Xianting and W.Lei, “Steganalysis using Regional Correlation and SecondOrder MarkovFeatures,” in *International Journal of Security and its Applications*, vol. 9, no. 1,2015, pp.69-76.

[16] J. Fridrich, T. Pevný and J. Kodovský, “ Statistically Undetectable JPEG Steganography:Dead ends, Challenges, and Opportunities,” in *Second InternationalWorkshop on Multimedia &Security,* 2007, pp.3–14.

[17] V. Holub and J. Fridrich,” Digital Image Steganography using Universal Distortion,” *in First ACM Workshop on Information Hiding and Multimedia Security*, 2013, pp.59–68.

[18] S. Bera and M. Sharma,” Frequency Domain Steganography System usingModified Quantization Table,” in *International Journal of Advanced and InnovativeResearch*, Vol.1, no. 7, 2012, pp.193-196.

[19] http://dde.binghamton,edu/download.edu/download/syndrome/.