©ELECTRON
Department of ECE, Amrita Vishwa Vidyapeetham, Coimbatore

FEATURE BASED CLASSIFICATION SYSTEM FOR JPEG STEGANALYSIS

Deepa D.Shankar, Gireeshkumar T., Jithin R.

TIFAC CORE in Cyber Security,Amrita Vishwa Vidyapeetham,Coimbatore,india

sudee99@gmail.com, gireeshkumart@gmail.com,jithin550@gmail.com

Abstract

The objective of steganalysis is to detect messages hidden in cover images, such as digital images. The ultimate goal of a steganalyst is to extract and decipher the secret message. In this paper, we present a powerful new blind steganalytic scheme that can reliably detect hidden data in JPEG images. This would increase the success rate of steganalysis by detecting data in transform as well as spatial domain. This scheme is feature based in the sense that features that are sensitive to embedding changes and being employed as means of steganalysis. The features are extracted in DCT domain. DCT domain features have extended DCT features and Markovian features merged together to eliminate the drawbacks of both. The blind steganalytic technique has a broad spectrum of analyzing different embedding techniques. The feature based steganalytic technique is used in the DCT domain to extract about 23 functionals and classify the dataset according to these functionals. The feature set can be increased to about 274 features by merging both DCT and Markovian features. The extracted features are being fed to a classifier which helps to distinguish between a cover and stego image. Support Vector Machine is used as classifier here.

Keywords: **Steganalysis, DCT, Principal Component Analysis, Feature Extraction, Support Vector Machine**

1. INTRODUCTION

Steganography is a means of communication in a covert manner so that anyone who inspects the message being exchanged cannot collect enough evidence to prove that the message has data hidden in it.

To mount an attack on a steganographic scheme, we need to show that it is possible to detect hidden data with a probability greater than random guessing. In this paper we propose a new steganalytic technique which can be applied to different steganographic schemes and image format. However it can be ideally used in JPEG format.

Steganalysis can be broadly classified as Blind Steganalysis and Targeted Steganalysis. Targeted Steganalysis are designed for a particular steganographic algorithm. This technique is more robust since it has good detection accuracy for that specific technique when they used against the particular steganographic technique. Blind Steganalysis are schemes which are independent of any specific embedding technique are used to alleviate the deficiency of targeted analyzers by removing their dependency on the behavior of individual embedding techniques. Hence one technique can work in a broad spectrum of steganographic techniques. This approach alleviates the deficiency of specific steganalysers by removing their dependency on the behavior of individual embedding techniques. To achieve this, a set of distinguishing statistics that are sensitive to a wide variety of embedding operations are determined and collected. These statistics, taken from both cover and stego images are used to train a classifier, which is subsequently used to distinguish between cover and stego images. Hence, the dependency on a specific embedder is removed at the cost of finding statistics that distinguish between stego and cover images accurately and classification techniques that are able to utilize these statistics.

Universal steganalysis is composed feature extraction and feature classification. In feature extraction, a set of distinguishing statistics are obtained from a data set of images. There is no well defined approach to obtaining these statistics, but often they are proposed by observing general image features that exhibit strong variation under embedding. The second component, feature classification, operates in two modes. First, the obtained distinguishing statistics from both stego and cover images are used to train a classifier. Second, the trained classifier is used to classify an input image as either being a clean image or carrying a hidden message.Previous literature [1] state only the application of JPEG images in the either DCT domain or in spatial domain [3] in terms of embedding and extraction.. Feature based steganalysis [1] [4]is a technique wherein certain features that are sensitive to embedding changes but insensitive to image content is extracted. A set of distinguishing features are obtained from DCT, DWT and spatial domains. This paper intends to merge both DCT and Markovian features [2] with a possibility of eliminating the drawbacks of both with a feature selection using Principal Component Analysis. The previous literature mentions about the high computational costs of other classifiers like non linear SVM. The feature selection is performed using Principal Component Analysis which can increase the computational complexity. In order to reduce further computational complexity or costs, and to obtain reasonable success, SVM is used as a classifier for the DCT domain. Fridrich [1] had developed a blind steganalytic scheme for feature based steganalysis.Fridrich et al [2] used standard 274 features by merging DCT and Markov features for JPEG steganalysis.

In the next section, we will discuss about the general architecture of the system. Section 3 will deal with the implementation issues regarding the architecture. The experimental results are discussed in section 4.Section 5 will have a short note on the future work.

2. IMPLEMENTATION ISSUES

Below is the system architecture of the feature based steganalytic system using DCT .The concept of feature extraction is combined with linear classification to devise an analytic system mainly for JPEG images.. It has been understood in literature survey that calculating the features directly in JPEG domain is more sensitive to a wider type of embedding algorithms. This is more evident if calibration is used. The direct calculation also enables a more straight forward interpretation of the influence of individual features on detection as well as easier formulation of design principles leading to more secure steganography.

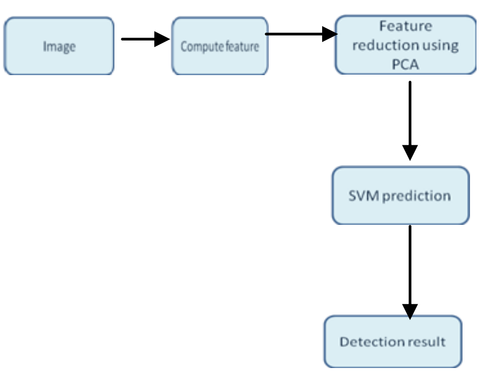

Fig1: System Architecture

2.1 FEATURE EXTRACTION

The goal of the paper is to merge new feature set which gives a better detection rate than any other steganalytic technique. The proposed feature set is used to construct a general linear classifier. The first step is to extract the features, and then Principal Component Analysis is used to find the optimal feature subset to improve the algorithm efficiency. Then the Support Vector Machine is designed with respect to accuracy, reliability and cost to give best results. The SVM is trained by the obtained features and then it is subjected to testing on images which were used during training and also on images which are not trained.JPEG format of images are considered here. Figure 1 shows the overall representation of the system.

Four types of features are extracted in DCT domain. They are First Order features, DCT features, Extended DCT features and Markov features. The first order statistics include, mean, standard deviation, skewness and kurtosis of pixels.DCT features are global histogram, individual histogram, dual histogram, variance, blockiness and cooccurance.The original DCT features [1] have 23 functional in them. The original DCT features can be extended to extended DCT features [2] which can extract about 193 functionals.Another set of features is the Markov features [2] whose dimensionality can be reduced to 81.While the extended DCT features model inter-block dependencies between DCT coefficients, the Markov features capture intra-block dependency among DCT coefficients of similar spatial frequencies within the same 8X8 block. Hence they have been merged to eliminate the drawbacks of both. Another reason for merging the set is that the classifiers employing each feature set individually have complementary performance. Markov features are unable to detect short message lengths. Hence feature sets are merged giving low false positive rates and high detection accuracy..

Messages less than 10% is not checked for detectability**.**All features in the merged feature set are calibrated. Calibration is a process through which one can estimate macroscopic properties of cover image from the stego image. Calibration is usually done in DCT domain which gives an estimate of cover image. It will be close to cover image in terms of perceptability.The image is converted to spatial domain, four pixels each are cut horizontally and vertically. This image is later converted back to DCT using the same quantization matrix. This is equivalent to shifting of am image, which helps it to retain the DCT coefficients .But the embedded data will be erased making the image a close estimate of the cover image. The newly obtained JPEG image *J2* has most macroscopic features similar to the original cover image because the cropped image is visually similar to the original image. The original DCT features were extracted as *L1* norm of the absolute value of the difference between the cover image and stego image. This process removes many relevant features needed for analysis. Hence certain functionals with proposed differences were used in DCT .They are called extended DCT features. The Markov feature set as proposed in [2] models the difference between absolute values of neighboring DCT coefficients as a Markov process. The Markovian functionals taken together will comprise of 324 features. This has increased dimensionality, which can be reduced by taking the average of the four 81 dimensionality features. Thus we get a combined set of Extended DCT and Markov features as 274.

2.2 FEATURE REDUCTION

PCA or Principal Component Analysis is used to analyze the effective dimensionality of DCT and spatial based feature space. This is used to reduce the dimensionality of features and hence to retrieve optimal dataset. The extracted dataset is fed into a PCA. Relevant features were extracted from a total set of data. These data have the maximum amount of information which helps to classify the cover images and stego image from a dataset that is inputted.

2.3 SVM PREDICTION

Support Vector Machine is a supervised learning technique for classification. There are many techniques for classification like neural network, perceptron, Fisher Linear Discriminant and SVM. Out of this, SVM is widely popular in Machine Learning since it maps the data from the original space into a high dimensional feature space. When the kernel function is linear, the resulting SVM is a maximum margin hyperplane.Given a training sample, the hyper plane splits a given training sample in such a way that the distance from the closest cases to the hyper plane is maximized. The complexity of SVM depends on the training samples. Hence we can conclude that SVM guarantees generalization to a great extend.

3. EXPERIMENTATION RESULT

3.1 DATABASE OF IMAGES

One of the important aspects of any performance evaluation work is the dataset employed in the experiments. The dataset needs to include a variety of textures, qualities and image formats. A set of 25 images were taken both from JPEG and BMP format and compressed to a size of 512 X 512. We choose a large amount of JPEG format due to its wide popularity in transmission through the internet. A practical evaluation of project is presented by testing unconditional steganalysis for two different algorithms with diverse embedding mechanism: S-TOOL and JP-Steg. Unless stated otherwise, all results were derived on samples from the testing set that were not used in any form during training. The JPEG and BMP formats are used for study. 25 datasets of cover image and stego image is taken for analysis. Out of this, 12 sets of data are of BMP format and the rest are in JPEG format. These images are used to extract different features like first order features, DCT features, Extended features etc.Many features of these datasets maybe irrelevant. Hence these features maybe removed for better performance. This dimensionality reduction may be achieved by using Principal Component Analysis. The output of Principal Component Analysis is given to a linear SVM. Out of the 25 datasets used, 17 are used to train the SVM.8 datasets are used for testing.The extracted features were checked without inputting in the PCA.The classification results were only 50%.

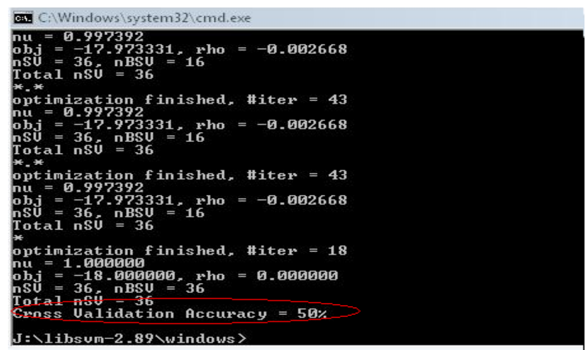

Fig 2: SVM classification without the use of PCA

Fig 2 shows the output of the SVM without implementing the PCA.The feature set dimensionality were later reduced using PCA. The result was 90% which is depicted in Fig 3.Thus the features have been reduced to give a more accurate classification.

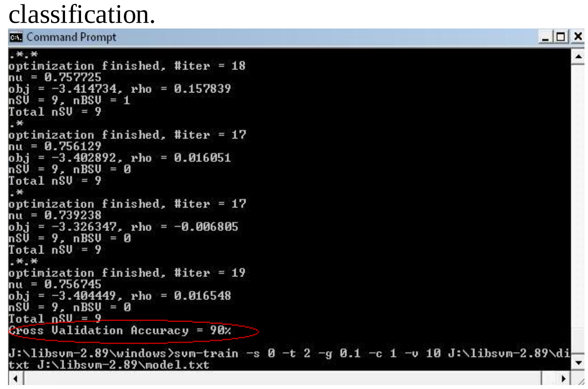

Fig 3: SVM Classification using PCA

3.2 PRINCIPAL COMPONENT ANALYSIS

A set of feature that have different statistics are extracted. The features are again reduced to obtain another set with lower feature dimensionality. This is only in terms of statistical outlook. The features can be further reduced by means of Principal Component Analysis. The first value found is called the principal component. The lower values are ignored after finding the Eigen values and Eigen vectors thus reducing the dimensionality. The Principal Component Analysis will reduce the dimensionality but may have a probability of eliminating the best features.

3.3 LINEAR SVM

The dimensionality reduction needs to be done because the classifier used here is linear SVM.There are many classifiers like neural network, perceptron, linear fisher discriminant, SVM etc. Out of these, SVM is considered to be more powerful in terms of classification. The feature reduction is mainly due to the use of linear SVM to reduce the cost and computational complexity. Since the steganalysis system used was Blind, the SVM has to be trained before any testing occurs. Out of the 25 images, 17 images were used to train the data and the rest 8 were used to test the data.

4. CONCLUSION AND FUTURE WORK

A set of features for steganalysis of JPEG with a range of quality factors was developed. We considered features that take into account the numerical changes in DCT coefficient introduced by embedding. The feature set was obtained by merging and modifying two previously proposed feature sets with complementary performance (the DCT features that capture the inter-block dependencies among DCT coefficients and Markov features which capture intra-block dependencies).In particular, we used the DCT features by replacing the L1 norm in their calibration by differences and added calibration to Markov features and reduced their dimensionality .According to the experiments, the new merged feature set provides better results than previous results.

The present feature based system has PCA incorporated for reduced dimensionality thereby obtaining a better feature set. This feature set is being input into the linear SVM for classification between cover and stego image. Apart from the features mentioned, it has been decided to find a set of calibrated set of 274 features and uncalibrated set of 274 features [5].feature selection can be done using independent component analysis. These features can be later fed to a classifier, probably a soft margin classifier with Gaussian kernel [5].The classification accuracy of the calibrated features as compared to uncalibrated features can be estimated .

Another enhancement of the paper can be an introduction to estimation of payload. This is called quantification, which can be achieved by means of Support Vector Regression [6].

The final enhancement is analyzing the LSB using autocorrelation between pixels. This is done with the assumption that the neighboring pixels are equal in value. The LSB planes matrix is formed. All 0's are replaced by -1 forming a new matrix. Every bit in the matrix is multiplied by itself and the results are all summed. If the two LSB s are equal, their multiplication is 1 else it is -1.So unequal bits decreases the correlation and equal bit increases. After extracting suitable separating features, weighted Euclidean distance measure are used .There are two initial sets of Stego and non-stego bit planes from which classifiers should be tuned. The centers and variance vectors of each set is computed and their Euclidean distance is found with the two centers. The results are known as auto correlation vector. The incoming vector is classified as stego if the Euclidean distance between the stego and center is less than the Euclidean distance.

REFERENCES

[1] J.Fridrich "*Feature Based Steganalysis for JPEG Images and its Implications for Future Design of Steganographic Schemes".*6th International Workshop, volume 3200 of lecture notes in Computer Science, Pages 67-81, 2005

[2] T.Pevn'y and J.Fridrich. "*Merging Markov and DCT Features for Multiclass JPEG Steganalysis"* Proceedings SPIE, Electronic Imaging, Security, Steganography and Watermarking of Multimedia Contents IX,Volume 6505, Pages 301-314, San Jose,CA,January 29-February 1, 2007

[3] Arezoo Yadollapour,Hossein Miar Niami " *Attack on LSB Steganography in Color and Grayscale Images using Autocorrelation Coefficients"* European Journal of Scientific Research,ISSN 1450-216 X Vol 31 No.2 (2009) Pages 172-183

[4] Mehdi Kharrazi,Husrev T.Sencar, Nasir Memon "*Performance study of common image steganography and steganalysis techniques"*Journal of Electronic Imaging15(4),041104(oct-Dec 2006)

[5] Jan Kodosky , Jessica Fridrich *"Calibration Revisited"*,ACM Multimedia and Security Workshop, Princeton, NJ, September 7–8, pp. 63–74, 2009.

[6] Tomas Pevny Jessica Fridrich, Andrew D.Ker "*From Blind to Quantitative Steganalysis"*, SPIE, Electronic Imaging, Media Forensics and Security XI, San Jose, CA, January 18-22, 2009, pp. 0C 1-0C 14.