VOL. 9, NO. 12, DECEMBER 2014 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences


www.arpnjournals.com

STEGANALYSIS WITH CLASSIFIER COMBINATIONS

J. Anita Christaline[1], R. Ramesh[2] and D. Vaishali[3]
[1, 2]Department of ECE, SRM University, Chennai, India
[2]Department of ECE, Saveetha Engineering College, Chennai, India
E-Mail: ani_chris@yahoo.co.uk

ABSTRACT

Blind steganalysis is based on choice of the feature set and the machine learning classifiers used for classification. While the performance of individual classifiers is good, the classification accuracy is seen to increase by appropriate combination of classifiers. This research has implemented image steganalysis with fusion of classifiers by various data fusion schemes. We intend to analyse the classification accuracy of fusion classifier under nine different fusion schemes. The chosen individual classifiers are Multi-Layer Perceptron (MLP) and Support Vector Machines (SVM). The feature set chosen for classification includes the calibrated, combined features of modified Discrete Cosine Transform (DCT) features and Markov features from a database of 1000 cover images and 1000 stego images. It has been identified that the classification accuracy of Decision template and Dempster-Shafer methods of fusion gives best results, while Bayes, average and sum fusion schemes give good results compared to the performance of individual classifiers.

Keywords: image steganalysis, fusion schemes, multi-layer perceptron, support vector machine, classification.

1. INTRODUCTION

Steganography (the art of covered writing) is the process of hiding secret information in digital media. In contrast to cryptography, the very existence of the information is hidden in steganography. Image steganography is an undetectable communication where the secret message is embedded into an innocuous digital image called cover. The embedded image is called as stego image. The embedding process is controlled by a stego key. While the main aim of steganography is to achieve statistically undetectable communication, its counterpart, steganalysis aims to break the steganographic systems by statistical measures of the images.

In recent past, blind steganalysis by classifying the feature vectors of the cover and stego images has gained momentum, both in JPEG and spatial domains. Good performance of steganalysis has been reported by Fridrich [1] when the features are directly derived from the cover image while embedding. The features chosen could also include the DCT coefficients[1] or the Markov features [2] or the global histograms or co-occurrence matrices [3] or a combination of these features [4]. The performance of the classifiers has been reported to improve by proper selection or calibration of the extracted features. Feature calibration was a method introduced by Fridrich, *et al*, [5], to estimate the histogram of the cover image with the knowledge of the histogram of the stego image. This concept was first used in attacking the F5 algorithm proposed by Westfeld [6].

Kodovsky and Fridrich [4] have analysed five different types of calibration according to the method of providing the feature and the corresponding reference value. They further propose an improved method of calibration which provides the reference value as additional feature rather than subtracting it from the original feature value. They claim that their new method removes failures encountered by normal methods and provides better steganalysis of varied steganographic schemes under different payload conditions.

With improvement in feature selection and calibration techniques for steganalysis [7], the need for implementing and fusing different machine learning classifiers [8] has become critical. Machine learning classifiers require the availability of best features as training data to arrive at an appropriate decision boundary. The performance of individual classifiers may differ for the same feature set. By combining the outputs of different classifiers, the accuracy of classification is said to be more than the accuracy of the best classifier. This combination is referred as fusion classifier or classifier pool or ensemble classifiers increase is based on data fusion schemes [9]. Prominent data fusion schemes proposed by Alkoot and Kittler [10] include median, average, minimum, maximum, majority voting. Apart from these, decision template, sum, Bayes, product, and Dempster-Schafer schemes can also be used for fusing the classifiers. In this paper we propose a combination of all these nine classifier fusion schemes for two different chosen classifiers, Support vector Machines (SVM) and Multi-Layer Perceptron (MLP). The performance of individual classifiers and the fusion classifiers in terms of accuracy has been studied.

2. EXPERIMENTAL SETUP

In this research work we intend to implement a fusion classifier schemethat could classify stego and cover images. The concept is implemented in MATLAB in three stages. The first stage is creation of stego images form cover images. The second stage is feature extraction part and the third stage is classification part.

A. Database of images used

The images used in this research include the raw. pgm images from the BOSS (Break Our Steganographic Systems) database. Standard set of cover images are

available as BOSSBase and BOSSRank [11]. The BOSSRank data base consists of 1000 grayscale cover images of size 512 x 512. These images were acquired with 7 different cameras and arranged in archives of 1000 images. These raw images have been taken and converted into the JPEG by MATLAB code. These 1000 cover images are used for embedding to create 1000 stego images. The total of 2000 images is used for training and testing of the chosen classifiers.

B. Chosen steganographic scheme

Statistics preserving steganographic schemes like OutGuess [12], Steghide, and Model Based Steganography have been proved to be vulnerable to statistics based steganalysis. The other category, heuristics based algorithms for steganography like F5, JP Hides and Seek, -F5 and nsF5 are more secured. For our experimental setup, we intend to use the nsF5 algorithm as the steganographic scheme. The logic behind nsF5 is that the embedding changes are done only on non-zero AC coefficients and the change is reflected as a decrease in the absolute value of the DCT coefficient by one. Shrinkage is said to occur when the coefficient becomes zero after embedding. In such situation, that particular coefficient is skipped and the next coefficient is used. This leads to lack of efficiency in terms of embedding capacity. The concept of wet paper codes [13] could be combined with F5 algorithm to improve the embedding capacity. This method called as nsF5 (no shrinkage F5) has been used as the steganographic scheme to generate stego images from BOSS cover images.

C. Selected image features

The selection of image features plays a vital role in the performance of the classifier. The first use of image feature set in terms of image quality metrics [14] to train classifier was based on binary similarity measures for steganalysis of LSB based steganographic systems. Later researches have presented features based on higher order moments [15]. For best steganalysis, the chosen features need to be sensitive to the embedding changes done during steganography and insensitive to the contents of the image itself. As the most common image format used is JPEG, our research intends to choose features that are compatible with the JPEG technique.

a) Markov random features

For JPEG images, Markov features that are based on the Markov models of the DCT plane have been defined [16]. The difference between adjacent absolute DCT coefficients is modelled as Markov process to get the Markov features. The Markov features are calculated from a base matrix I(x, y) that contains the DCT coefficients in 8 x 8 blocks. From this base matrix, four difference arrays (horizontal, vertical, diagonal and diagonal minor) are calculated as follows,

$I_h(x, y) = I(x, y) - I(x + 1, y)$,
$I_v(x, y) = I(x, y) - I(x, y + 1)$,
$I_d(x, y) = I(x, y) - I(x + 1, y + 1)$,
$I_m(x, y) = I(x + 1, y) - I(x, y + 1)$.

The features are calculated from four different transition probability matrices (Mh, Mv, Md, Mm) that are formed from the above difference matrices. The elements of these when matrices taken as such would lead to a feature vector with large dimensionality, hence the dimensionality can be brought down by choosing a window of 9 x 9 in each of the four matrices. This would give a dimension of 4 x 81 = 324, which can further be reduced by considering the average of all the four matrix value. This would give the feature dimension as 81. Further, the values of this feature set may be normalised before finding the transition matrices.

b) DCT features

As most of the steganographic techniques use the DCT coefficients for embedding, these DCT features could be the best parameters for defining the features of the image. Considering the DCT coefficients of the luminance, one of the modified features set could be the histogram of the 64 DCT coefficients giving 11 features. The feature set is further extended by including the AC histograms (5 features), dual histograms (99 features), co-occurrence matrix (25 features), blockiness (2 features) along with the variation[3]. This extended feature set now has 193 features. Thus the combined feature set with the extended DCT and Markov feature has 193 + 81 = 274 features.

D. Feature set calibration

Calibration of the feature set recommended by many researchers adopts different methods [1], [13], [17]. While the general method for calibration is finding the feature difference between the original image and the reference image, the Cartesian product of the feature sets of the two images gives better performance while classification. The reference image is created by converting the JPEG image to spatial domain and then cropping. After cropping it is converted back to the JPEG format. The algorithm for this calibration is as follows,
Algorithm for Calibration

a) Read JPEG image
b) Get the features of this original image
c) Apply Inverse DCT (IDCT) to get image in spatial domain
d) Crop the image with suitable size
e) Apply DCT to get reference image
f) Get the features of this reference images
g) Perform Cartesian product of the features of the reference image and the original image.

The combined feature set with 274 features is subjected to calibration to get another 274 features. The overall feature set is framed by adding the original feature

set and the calibrated feature set. This feature set has 548 features in total.

3. RESEARCH FRAMEWORK FOR CLASSIFIER FUSION

A. Machine learning classifiers

Machine learning classifiers depend on the availability of large data sets for training and validation. The popular supervised machine learning classifiers [9] are the multilayer perceptron classifier (MLP), the support vector machines classifier (SVM), the decision tree (DT) classifier, *K*–Nearest Neighbourhood (KNN). While many steganalysis research has considered support vector machines classifier (SVM) as the base classifier, little research has been conducted in combining multilayer perceptron classifier (MLP) and the support vector machines classifier (SVM). Hence we propose to implement a fusion classifier with MLP and SVM and compare the performance of individual and combined classification accuracies.

a) Chosen classifiers

For our research, we intend to choose the MLP and SVM classifiers. The multilayer perceptron (MLP) is a popular form of neural network for various pattern recognition and steganalysis applications. MLP has the basic architecture illustrated in Figure-1. The network takes the feature values and combines them with suitable weight vectors that are treated as inputs. The output is based on the networks activation function. Research by Windeatt [18] states, that MLP is powerful classifiers with many free parameters. The weights for MLPs are trained with the back propagation algorithm such that they can associate a high output response with particular input patterns.

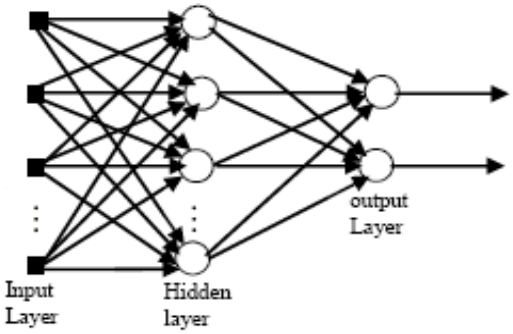

Figure-1. MLP architecture with one hidden layer.

Support Vector Machine (SVM) is a powerful tool for steganalysis and pattern classification [7] and has been widely used by many researchers. SVM uses nonlinear mapping to map the input vectors unto a high dimensional feature space where the decision boundary for linear classification is constructed. Figure-2 presents a sample hyper plane of a SVM classifier for the linear separable case.

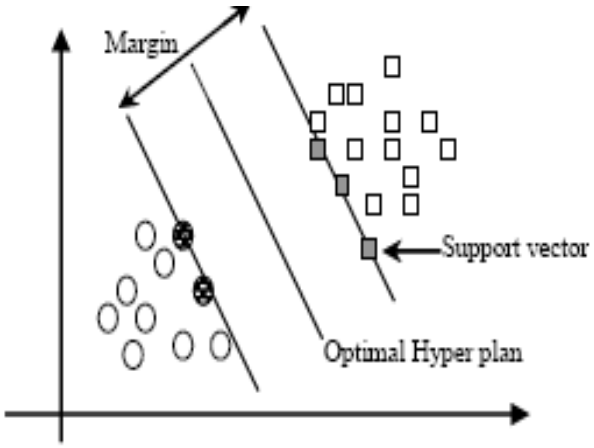

Figure-2. SVM classifier hyperplane for linear separable case.

When the distance between the two data classes is more, the separation is good. The support vectors are the samples around the margin. The successes of SVM depend on the kernel (inner product) during mapping.

B. Fusion techniques

Though individual classifiers perform well, their performance is said to increase by fusion methods. Ludmila [19] presents six methods of combining the classifiers. Apart from these, decision template, sum, Bayes, product, and Dempster- Schafer schemes can also be used for fusing the classifiers. In this paper we propose a combination of all these nine classifier fusion schemes. The frame work for fusing the chosen classifiers in our research is presented in Figure-3.

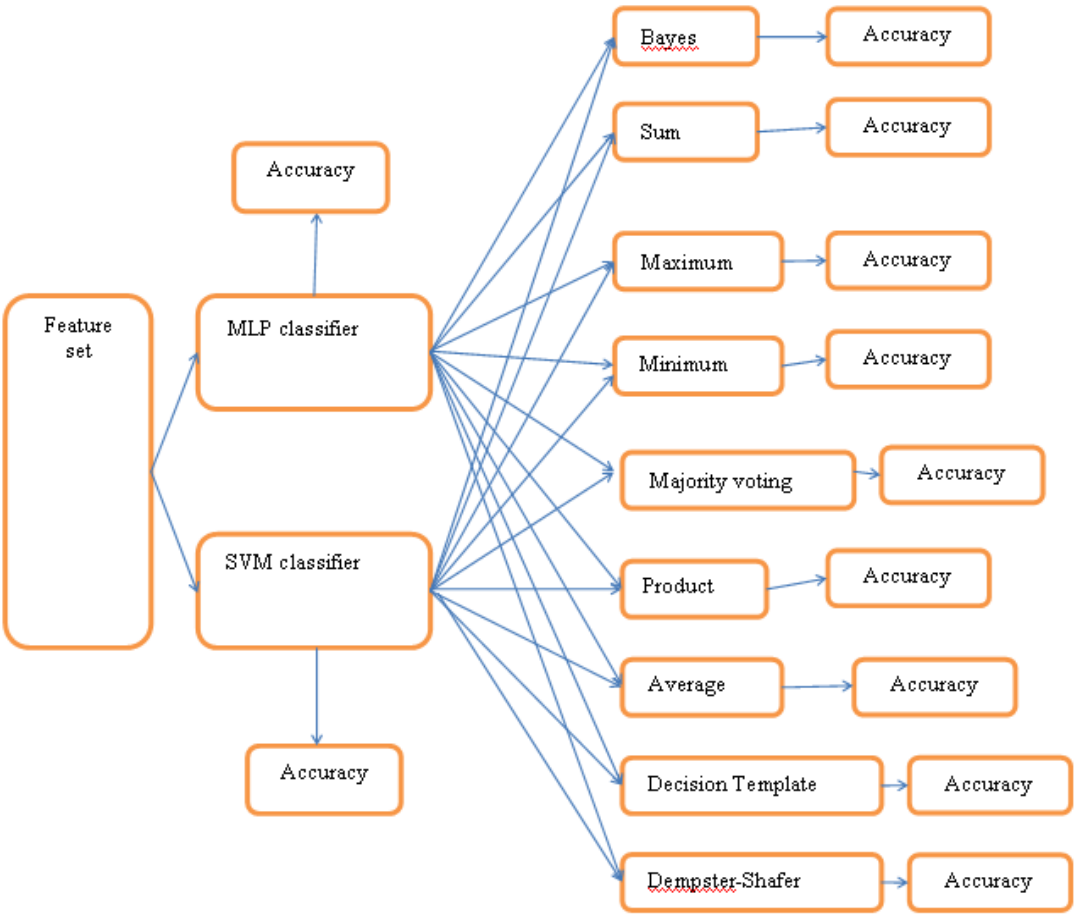

Figure-3. The frame work for fusing the chosen classifiers.

4. ANALYSIS OF DATA

We implemented our research in three stages using MATLAB. In the first stage, the 1000 cover images are subjected to nsF5 steganographic code to create 1000 stego images. In the second stage, we extracted the features as described in section 2 above. These features are represented by two matrices (one is cover feature set and another is stego feature set) each of dimension 1000 x 548.In the third stage, these feature sets are used for training, testing and validating the classifiers individually and with fusion techniques.

A. Performance of classifiers

The performance metric chosen for comparing the individual classifiers and the fusion classifiers is the 'Accuracy' in classification. Accuracy is the percentage of correct predictions. Numerically it is the calculation of , Accuracy = (TP + TN) / (TP + TN + FP + FN), where TP is number of True Positive, TN is number of True Negative, FP is number of false Positive and FN is number of False Negative. Though there are other metrics like precision, sensitivity and specificity for measuring the performance of classifiers, accuracy seems to be the most important metric as it considers all possibilities of positive and negative outputs of classifiers.

The individual classifiers chosen by us include MLP and SVM. The accuracy of these is shown in Table-1.

Table-1. Accuracy Analysis of Individual Classifiers.

Classifier type	Accuracy
MLP	0.3533
SVM	0.6111

From the Table it has been seen that the accuracy of SVM classifier is good compared to MLP.

The accuracy of the fusion classifier (combined accuracy of MLP and SVM) with different fusion schemes is shown in Table-2.

Table-2. Accuracy analysis of fusion classifiers under different fusion schemes.

Fusion scheme	Accuracy
Majority voting	0.6111
Maximum	0.6111
Sum	0.6667
Min	0.3533
Average	0.6667
Product	0.6111
Bayes	0.6667
Decision Template	0.8900
Dempster-Shafer	0.9800

5. DISCUSSIONS OF RESULTS

Best level of accuracy (0.9800) has been noticed for Dempster-Shafer fusion scheme, followed by Decision template (0.8900). Good performance (0.6667) has been noticed with the fusion schemes - Bayes, average and sum. These five schemes give accuracies that are greater than the individual classifiers. Average accuracy (0.6111) is noticed in product, maximum and majority voting schemes. The least accuracy (0.333) is that of minimum fusion scheme. Thus, Decision template, Dempster-Shafer, Bayes, average and sum fusion schemes yield better classification in image steganalysis for the chosen feature set.

6. CONCLUSIONS

In our research we have implemented image steganalysis with fusion classifiers. We have analysed the classification accuracy of fusion classifier under nine different fusion schemes. The chosen classifiers are Multi-Layer Perceptron (MLP) and Support Vector Machines (SVM). The feature set used for classification includes the calibrated, combined features of modified DCT features and Markov features. This feature set is seen to provide better classification from the database of 1000 cover images and 1000 stego images. The classification accuracy of Decision template and Dempster-Shafer methods of fusion gives best results, while Bayes, average and sum fusion schemes give good results compared to the performance of individual classifiers.

REFERENCES

[1] Fridrich J. 2005. Feature-based steganalysis for JPEG images and its implications for future design of steganographic schemes. In Information Hiding, 6th International Workshop. 3200: 67-81.

[2] Shi YQ, Chen C, Chen W. 2006. A Markov process based approach to effective attacking JPEG steganography. In Proceedings of the 8-th Information Hiding Workshop.

[3] Pevny T, Fridrich J. 2007. Merging Markov and DCT features for multi- class JPEG steganalysis. In: E. J. Delp and P. W. Wong, editors. Proceedings SPIE, Electronic Imaging, Security, Steganography, and Watermarking of Multimedia. 6505: 3-14.

[4] Kodovsky J, Fridrich J. 2009. Calibration Revisited. in J. Dittmann, S. Craver and J. Fridrich, editors, Proceedings of the 11th ACM Multimedia and Security Workshop.

[5] Fridrich J, Goljan M, Hogea D. 2002. Steganalysis of JPEG images: Breaking the F5 algorithm. In Information Hiding, 5th International Workshop. 2578: 310-323.

[6] Westfeld. 2001. High capacity despite better steganalysis (F5 - a steganographic algorithm). In: I. S. Moskowitz, editor, Information Hiding, 4th International Workshop. 2137: 289-302.

[7] Kodovsky J, Fridrich J, Holub V. 2012. Ensemble Classifiers for Steganalysis of Digital Media. IEEE Transactions on Information Forensics and Security. 7(2): 432-444.

[8] Parikh D, Polikar R. 2005. A Multiple Classifier Approach for Multisensor Data Fusion. Proceedings of Ieee Fusion. 1: 453-460.

[9] Boujelbene SZ, Dorra BA, Noureddine E. 2011. General Machine Learning Classifiers and Data Fusion Schemes for Efficient Speaker Recognition. International Journal of Computer Science and Emerging Technologies. 2(2).

[10] Alkoot F, Kittler J. 1999. Experimental Evaluation of Expert Fusion Strategies. Pattern Recognition Letters. 20: 1361-1369.

[11] Bas P, Filler T, Pevny T. 2013. Break Our Steganographic System --- the ins and outs of organizing BOSS. In: proceedings of Information Hiding Conference 2011; (): http://exile.felk.cvut.cz/boss/BOSSFinal/index.php?mode=VIEW&tmpl=materials Accessed (accessed 15 November).

[12] Fridrich J, Goljan M, Hogea D, Soukal D. 2003. Quantitative steganalysis of digital images: Estimating the secret message length. ACM Multimedia Systems Journal. 9(3): 288-302.

[13] Fridrich J, Goljan M, Soukal D. 2006. Wet paper codes with improved embedding efficiency. IEEE Transactions on Information Security and Forensics. 1(1): 102-110.

[14] Avcibas I, Memon N, Sankur B. 2001. Steganalysis using image quality metrics. In E. Delp and P. W. Wong, editors, Proceedings of SPIE Electronic Imaging, Security and Watermarking of Multimedia. 4314: 523-531.

[15] Farid H, Lyu S. 2006. Steganalysis using higher-order image statistics. IEEE Transactions on Information Forensics and Security. 1(1): 111-119.

[16] Shi YQ, Chen C, Chen W. 2006. A Markov process based approach to effective attacking JPEG steganography. Proceedings of the 8th Information Hiding Workshop.

[17] Provos N. 2001. Defending against statistical steganalysis. 10th USENIX Security Symposium. 323-335.

[18] Windeatt T. 2006. Accuracy/diversity and ensemble MLP classifier design. IEEE TRANSACTIONS ON NEURAL NETWORKS. 17(5): 1194-1211.

[19] Ludmila IK. 2002. A Theoretical Study onSix Classifier Fusion Strategies. Ieee Transactions on Pattern Analysis and Machine Intelligence. 24(2).