4-2007

Multi-Class Classification Averaging Fusion for Detecting Steganography

Benjamin M. Rodriguez
Air Force Institute of Technology

Gilbert L. Peterson
Air Force Institute of Technology

Sos S. Agaian
University of Texas at San Antonio

Recommended Citation

B. M. Rodriguez, G. L. Peterson and S. S. Agaian, "Multi-Class Classification Averaging Fusion for Detecting Steganography," 2007 IEEE International Conference on System of Systems Engineering, San Antonio, TX, USA, 2007, pp. 1-5, doi: 10.1109/SYSOSE.2007.4304292.

Multi-Class Classification Averaging Fusion for Detecting Steganography

Benjamin M. Rodriguez
Dept of Elect and Computer Engr
Air Force Institute of Technology
Wright Patterson, AFB, OH, USA
benjamin.rodriguez@afit.edu

Gilbert L. Peterson
Dept of Elect and Computer Engr
Air Force Institute of Technology
Wright Patterson, AFB, OH, USA
gilbert.peterson@afit.edu

Sos S. Agaian
Multimedia & Mobile Sig Lab
Dept of Elect and Computer Engr
The Univ of TX at San Antonio
sos.agaian@utsa.edu

Abstract - *Multiple classifier fusion has the potential to more accuratly perform classification than each of the individual classifiers alone. Classifier fusion is often based on fixed combination rules like the product and average rules. In the literature, multiple classifier fusion systems have proved to be a valuable approach to combining classifiers. In classifier selection the classifiers are picked from a larger pool of classifiers. This paper focuses on multi-class classification fusion based on weighted average of posterior class probabilities. The motivation of this fusion system is on identifying the stego fingerprint within jpeg images. The embedding methods targeted are F5, JSteg, Model Based, OutGuess, and StegHide. The embedding methods used present different challenges when attempting to extract the hidden information. This challenge is due to changes caused by embed data in dramatically different ways within the jpeg image. The classifier ensemble used in this approach is selected based on the individual performances of each classifier. These classifiers include kernel based classification systems, SVM, Kernel Fishers discriminant along with Expectation Maximization and Parzen Windows. The system consists of three levels: feature preprocessing, classifier system, and fusion.*

Keywords: Fusion System, Multi-class Classification, Steganography, Steganalysis

1 Introduction

In steganography the primary goal is to hide a hidden message from being seen by an outside observer. If an embedded message is discovered the primary goal of steganography is defeated, so concealing the existence of a hidden message is essential. Steganalysis on the other hand tries to identify a file as containing hidden information or not. Identifying the embedding method is an important step in digital forensics prior to extracting the data hidden through steganography in an image. The identification of the embedding technique consists of performing multi-class classification using known embedding steganography signatures. The identification of the embedding methods focuses on multi-class identification. In this paper targeted detection is used with an objective to determine patterns that result from hiding the message left behind by the steganography method. This is referred to as a stego fingerprint. Finding the stego fingerprint is essential for an analyst if the hidden information is to be extracted.

Steganalysis has several directions in which the systems are designed to determine if hidden information has been embedded within a digital image. Methods that are based on classifying extracted features focus on simple classification method. Although decrypting signals has been around for several decades, steganalysis is still a relatively new science. Higher order statistic and wavelet based stego detection method was presented by Lyu and Farid [7] using Fisher's linear discriminant and support vector machines. A multilevel based feature method with a wavelet structure was developed by Agaian, et al. in 2004 [1] which focus' on localization of stego information within DCT 8 by 8 blocks and 16 by 16 blocks. These methods were not designed as a multi-class classification system but rather targeting system to identify individual embedding method individually. Rodriguez and Peterson in [9] presented a multi-class classification system which focused on identifying the stego fingerprint within jpeg images.

While other methods have been developed for multi-class classification a fusion system is needed to improve the classification accuracy of individual classifiers. The main focus of recent research in classifier fusion has been on establishing the relationship between the diversity of the classifiers and their resulting accuracy/performance. To combine classifier systems together, the newly fused system should perform better than the individual systems and possibly other fused systems. Ideally, the performance of the fused system should be appropriately measured, accounting for all dependency between the individual systems so that the performance is not over estimated or under estimated by assumptions of independence.

This paper presents a multi-class detection method aimed at identifying steganography embedding methods used to hide digital data within cover jpeg images. The goal is to implement a multi-class fusion system for classification of steganographic methods. The fusion technique is to be used for multi-class classification of steganography techniques. Improving the classification accuracy in a multi-class classification system can be accomplished with classifier fusion. A weighted fusion system is proposed which performs better than individual multi-class classifier or simple non weighted system. The basic structure of the fusion technique proposed is a three level classification system. The first level is a two-class preprocessing system in which the input features are preprocessed and mapped into a new feature space with the intent of creating a larger separation between classes. In [8] Rodriguez et. al proposed a feature selection (classification dependent) in the kernel space to determine which features are relevant to the targeted embedding method which is used here during the preprocessing stage. In classifier selection the classifiers are picked from a pool of classifiers are used based on the performance of the possible permutations and the performance of the selected classifiers. This is the second level of the proposed system, classifier system. The outputs of each classifier are posterior class probabilities, the range of the classifier outputs are [0 1], intermediate feature space. The intermediate features are used as the input into the classifier fusion. Each vector in this set is an expanded version of the intermediate features used as the inputs into the final level of system, classifier fusion. The decision of this system is assigned as the most likely class label assignment based on the proposed weighted mean. The decision template for class label *i* is the average (the decision within this fusion could also be min, max, median, majority, etc.) of the system for the elements of the training data set labeled in class *i*. The higher the similarity between the system values of the current input feature $\hat{\mathrm{x}}$ and the class *i* the higher the validation for that class to be assigned. The proposed system is shown in Figure 1 where the three levels are enclosed with doted lines.

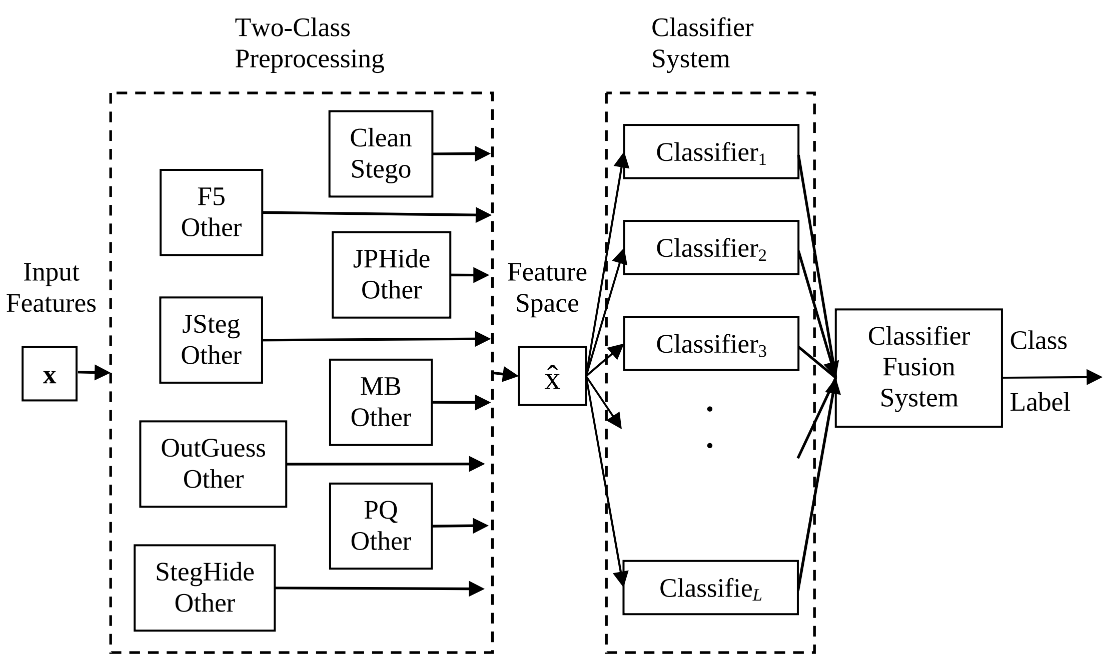

Figure 1 Classifier Fusion System

The classifier fusion system is discussed in section 2 along with simple fusion combiners. In section 3 the results from the proposed system are presented. A conclusion is presented in the last section.

2 Classifier Fusion System

This section presents simple classifier fusion combiners and weighted system. In each of these methods calculation for the support of class ω_j using only the j^{th} column of the decision system is defined by $\mu_j(\mathbf{x}) = \mathcal{F}\left[d_{1,j}(\mathbf{x}),...,d_{L,j}(\mathbf{x})\right]$, where $\mathcal{F}$ is a combination function and $d_{i,j}(\mathbf{x})$ is the set of class labels of classifier D_i which gives to the hypothesis that **x** comes from class ω_j. The class label of **x** is found as the index of the maximum $\mu_j(\mathbf{x})$. The combination function $\mathcal{F}$ can be chosen in many different ways. Weighted systems assign weights, w_i, to the most important classifier. The weights are generated with a variety of methods.

2.1 Average Fusion Combiners

Some popular choices of simple averaging classifiers are listed in Table 1. The class labels represented by $d_{i,j}(\mathbf{x})$ from the set of classifiers is considered, where $i = 1,\ldots,L$, as L point estimates of the same uncertain quality $P(\omega_j|\mathbf{x})$.

Table 1. Window Function Characteristics

Averaging	**Average Fusion Functions**
Simple	$\mu_j(\mathbf{x}) = \frac{1}{L}\sum_{i=1}^{L} d_{i,j}(\mathbf{x})$
Trimmed	$\mu_k(\mathbf{x}) = \frac{1}{R}\left(x_{k=1} + x_{k+2} + \cdots + x_{n+k}\right)$
Harmonic	$\mu_j(\mathbf{x}) = \left(\frac{1}{L}\sum_{i=1}^{L}\frac{1}{d_{i,j}(\mathbf{x})}\right)^{-1}$
Geometric	$\mu_j(\mathbf{x}) = \left(\prod_{i=1}^{L} d_{i,j}(\mathbf{x})\right)^{\frac{1}{L}}$
Weighted	$\mu_j(\mathbf{x}) = \sum_{i=1}^{L} w_i d_{i,j}(\mathbf{x})$

The trimmed mean uses K percent trimmed values. The L degrees of support are sorted and K percent of the values are dropped on each side. The overall support $\mu_j(\mathbf{x})$ is found as the mean of the remaining degrees of support. For example, given a set of observations, $\mathbf{x}$ 1) find n = number of observations 2) reorder the data as "order statistics" x_i in ascending order 3) find lower case $p = P/100$ = proportion trimmed 4) compute $n{\cdot}p$. If $n{\cdot}p$ is an integer use $k= n{\cdot}p$ and trim k observations at both ends. R = remaining observations = $n - 2k$. The geometric mean is equivalent to the product combiner as raised to the power of $1/L$ is a monotone transformation that does not depend on the class label j and therefore will not change the order of $\mu_j(\mathbf{x})$: the winning label obtained from the product combiner will be the same as the winning label from the geometric combiner. Table 2 shows the NAME weighted classifier considered in this paper. The estimate $\mu_j(\mathbf{x})$ of $P(\omega_j|\mathbf{x})$ is calculated by taking the weighted average and restricting the coefficients w_i to sum up to one, $\Sigma\ w_i = 1$. This weighted system contains L weights, one weight per classifier. The weight for a classifier D_i is based on its estimated error rate.

2.3 Weight Selection

In [6] the weights are derived so that they minimize the variance of $\mu_j(\mathbf{x})$. Since it is assumed that the estimators are unbiased the variance of each of the estimates $d_{i,j}(\mathbf{x})$, $i = 1,\ldots,L$, is equivalent to its expected squared error.

Table 2. Weighted Function Characteristics

Weight Function	**Weight Selection** w_i
Bartlett	$\left(\frac{L-i+1}{\sum_{i=1}^{L} L-i+1}\right)$
Hanning	$\frac{\left(1+\cos\left(\frac{2\pi i}{2(L+1)}\right)\right)}{\sum_{k=1}^{L}\left(1+\cos\left(\frac{2\pi k}{2(L+1)}\right)\right)}$
Hamming	$\frac{0.54+0.46\cos\left(\frac{2\pi i}{2(L+1)}\right)}{\sum_{k=1}^{L}\left(0.54+0.46\cos\left(\frac{2\pi k}{2(L+1)}\right)\right)}$
Blackman	$\frac{0.42+0.5\cos\left(\frac{2\pi i}{2(L+1)}\right)+\left(\frac{0.08(2\pi i)}{2(L+1)}\right)}{\sum_{k=1}^{L} 0.42+0.5\cos\left(\frac{2\pi k}{2(L+1)}\right)+\left(\frac{0.08(2\pi k)}{2(L+1)}\right)}$
Gaussian	$\frac{\exp\left(-\frac{1}{2}\left(\frac{i}{L+1}\right)^2\right)}{\sum_{k=1}^{L}\exp\left(-\frac{1}{2}\left(\frac{k}{L+1}\right)^2\right)}$
Kuncheva	$\frac{\frac{1}{\sigma_i^2}}{\sum_{k=1}^{L}\frac{1}{\sigma_k^2}}$

One version used by Kuncheva in [6] is a constrained regression for finding weights that minimize the variance of L weights is derived by assuming that the expert's errors in approximating the posterior probability, $P(\omega_j|\mathbf{x}) - d_{i,j}(\mathbf{x})$, are normally distributed with zero mean. Denoted by σ_{ik} the covariance between the approximation errors by classifiers D_i and D_k. $\mathbf{w} = \Sigma^{-1}\mathbf{I}(\mathbf{I}^{\ T}\ \Sigma^{-1}\mathbf{I})$ where $\mathbf{w} = [w_1,\ldots,w_L]^T$ is the vector of weights, Σ is the covariance matrix for the classifiers' approximation error and $\mathbf{I}$ is an L-element vector with ones. Assuming that the classifier output for class ω_j were independent. Then Σ in the equation for $\mathbf{w}$ is diagonal with the variance of $D_1,\ldots,D_L$ along the diagonal. In this case the weights are proportional to the inverse of the variances $w_i \propto (1/\sigma_i^2)$, and the equation for $\mathbf{w}$ reduces to the function in Table 2 labeled Kuncheva. For classification purposes the value of $\mu_j(\mathbf{x})$ does not have to be an unbiased estimate of the posterior probability, therefore the weights do not have to be constrained by the coefficients w_i to summing up to one. It can be said that the larger the weight the more important the classifier.

While the previous method assigns weights based on most important classifiers, this method may assign moderate weights to classifiers which may have 50% classification accuracy. This is due to the calculation of the variance between each of the classifiers if a large number of classifiers create a cluster near 50% accuracy. A simple set of weight assignments is proposed that assigns weights in descending order based on weight function. This will allow the classifiers with the largest classification accuracy to be assign the largest weights and the classifiers with the worst classification accuracy to be assign the smallest weight values. Various functions have been theoretically derived for a variety of needs. The characteristics for various proposed weighted function are listed below.

3 Results

A comparison between the individual classifiers and the fused classification strategies demonstrates the performance of each technique on the steganalysis problem. The images used consist of 1000 512 by 512 RGB jpeg images which consist of clean images sets and steganography images altered with the five embedding methods. The amount of hidden information embedded within each of the files was 4000 characters which is equivalent to one page of text. The features are generated from the energy band of the DCT coefficients within 8 by 8 blocks of the jpeg image.

Table 3. Classification Accuracy for Weight Selection

Fusion Method	**No. of Classes**	**Performance**
No Fusion - SVM	6	72.0%
Average System	6	84.4%
Proposed Weights	6	91.7%
Kuncheva Weights	6	92.0%

The comparison is made between simple multi-class classification and fusion techniques as shown in Table 3. The combination of a preprocessing level and the use of a weighted fusion system increase the classification accuracy over simple averaging fusion systems. While the proposed system is slightly outperformed by the Kuncheva method [6] the weight calculations for the proposed system are simple to calculate.

Conclusion

In this paper a weighted fusion system was presented in which the system determines the steganographic method used to create a stego image. The weighted fusion method is developed to assign weights to the classifier with the most importance when assigning the class label. The increase in classification accuracy of the presented classifier fusion system is an important step in identifying the stego method of the steganalysis system. This fusion system is easily adapted for other machine learning problems.

References:

[1] S.Agaian, "Color Wavelet Based Universal Blind Steganalysis", The 2004 International Workshop on Spectral Methods and Multirate Signal Processing, SMMSP 2004.

[2] Cabrera, J. B. D., "On the impact of fusion strategies on classification errors for large ensembles of classifiers", Pattern Recognition 39 (2006) 1963 – 1978

[3] Goebel, K. and Yan, W., "Choosing Classifiers for Decision Fusion"

[4] Grim, J., Kittler, J., Pudil, P. and Somol, P., "Multiple Classifier Fusion in Probabilistic Neural Networks", Pattern Analysis & Applications (2002)5:221–233

[5] Kittler, J., "A Framework for Classifier Fusion: Is It Still Needed?**",** Lecture Notes In Computer Science; Vol. 1876**,** Proceedings of the Joint IAPR International Workshops on Advances in Pattern Recognition, pp. 45 – 56, 2002, Springer-Verlag London, UK

[6] Kuncheva, L. I., *Combining Pattern Classifiers Methods and Algorithms*, Copyright © 2004 John Wiley & Sons, Inc.

[7] Lyu, S. and Farid, H., "Steganalysis Using Higher-Order Image Statistics", IEEE Transactions on Information Forensics and Security, 1(1):111-119, 2006

[8] Rodriguez, B., Peterson, G. and Bauer, K., "Steganalysis using SVM and Feature Saliency"

[9] Rodriguez, B. and Peterson, G., "Steganography Detection Using Multi Class Classification", IFIP WG11_9 Digital Forensics 2007.