# Detection of Stego Anomalies in Images Exploiting the Content Independent Statistical Footprints of the Steganograms

S. Geetha , Siva S. Sivatha Sindhu
Faculty, Department of Information Technology
Thiagarajar College of Engineering
Madurai-625 015, Tamil Nadu, India.
E-mail: sgeetha@tce.edu

N. Kamaraj
Department of Electrical and Electronics Engineering
Thiagarajar College of Engineering
Madurai-625 015, Tamil Nadu, India.
E-mail: nkeee@tce.edu

**Keywords:** image steganalysis, content independent distortion measures, genetic-X-means classifier

**Received:** September 1, 2008

*Steganography which facilitates covert communication creates a potential problem when misused for planning criminal activities. Its counter measure steganalysis is focused on detecting (the main goal of this research), tracking, extracting, and modifying secret messages transmitted through a subliminal channel. In this paper, a feature classification technique, based on the analysis of content independent statistical properties, is proposed to blindly (i.e., without knowledge of the steganographic schemes) determine the existence of hidden messages in an image. To be effective in class separation, the genetic-X-means classifier was exploited. For performance evaluation, a database composed of 5600 plain and stego images (generated by using seven different embedding schemes) was established. Based on this database, extensive experiments were conducted to prove the feasibility and diversity of our proposed system. Our main results and findings are as follows:*

1. *a 80%+ positive-detection rate.(promising rate for a blind steganalyzer)*
2. *The removal of content dependency from features enhances the discriminatory power of the classifier.*
3. *Universal, blind steganalyzer. (not limited to the detection of a particular steganographic scheme )*
4. *Detection of stego images with an embedding rate as low as 5% of the maximum payload.*

*Povzetek: Opisana je metoda iskanj skritih sporočil v slikah.*

## 1 Introduction

Steganography has been known and used for a very long time, as a way to establish covert communication between parties, by embedding the secret message in another, apparently innocuous, document. The goal of steganography is to communicate as many bits as possible without creating any detectable artifacts in the cover-object. Although steganography is an ancient subject, its modern formulation is often given in terms of the *prisoner's problem* by Simmons in 1983 [1]. In today's digital world, this has taken a new facet, however, and it must be approached in a spanking new view.

Due to the proliferation of the digital media and the easy accessibility to Internet, development of new technologies for network based multimedia systems and advanced multimedia services have been intensified. Many of the multimedia processing operations like editing, storage, transmission, and access of multimedia are easily done by any subject. Early methods exploited cryptography for secure transmission, to prevent unauthorized access and tampering of secret messages. However, the encrypted form may attract special attention of network warders and is thus not fully secret. Current information hiding techniques are developed to deceive warders by embedding messages into multimedia in an imperceptible manner, but still maintain their original formats and quality. Unlike cryptography, where the goal is to secure communications from an eavesdropper, steganographic techniques strive to hide the very presence of the message itself from an observer.

Steganography may provoke negative effects in the outlook of personal privacy, business activity, and national security. The scandalous can abuse the technique for planning criminal activities. For example, commercial spies or traitors may thieve confidential trading or technical messages and deliver them to competitors for a great benefit by using hiding techniques. Terrorists may also use related techniques to cooperate for international attacks (like the 9/11 event in the U.S.) and prevent themselves from being traced. Some others may even think of the possibility of conveying a computer virus or Trojan horse programs via data hiding techniques. Thus, it raises the concerns of enhancing warders' capability and lessening these

negative effects by developing the techniques of "steganalysis".

It should be noted that the primary goal of steganography is to set up a subliminal communication channel in a completely undetectable manner. In this context, "*steganalysis*" refers to the set of techniques that are designed to distinguish between cover-objects and stego-objects. Even though nothing might be gleaned about the contents of the secret message, when the existence of hidden message is known, revealing its content is not always necessary. Just disabling and rendering it useless will defeat the very purpose of steganography. This implies that the warder should be capable of discriminating suspicious objects from a large number of innocuous ones (i.e., the so-called passive steganalysis [3], [5]). In contrast to passive steganalysis, the goal of active steganalysis is to retrieve, modify, and even fabricate the embedded messages for destroying or interfering with covert communications and rendering hidden data useless. Applications of steganalysis then include, for example, an inlet/outlet content-monitoring program that inspects and intercepts suspected multimedia data transmitted on the network. In addition, steganalysis techniques can also be utilized to evaluate the security of covert communication channels under construction.

On the outset, deciding whether the cover media contains any secret message embedded in it or not is essential to steganalysis. Although it is uncomplicated to inspect suspicious objects and extract hidden messages by comparing them to the original versions, the restricted portability and accessibility of original cover-signals generally make blind steganalysis more attractive and reasonable in many practical applications. Blindness is meant to analyze stego-data without knowledge of the original signal and without exploiting the embedding algorithm. Hence, detecting the existence of hidden information becomes quite difficult and complex without exactly knowing which embedding algorithm, hiding domain, and steganographic keys were used. Apart from these issues, a steganalysis algorithm is required to possess other properties such as low complexity and low classification risk. A low-complexity algorithm makes the system capable of inspecting objects at a high throughput rate. An algorithm of low classification risk generally makes tradeoffs between costs resulting from missing errors (i.e., false negative) and from false alarms (i.e., false positive). This motivates our current research: devising a content independent feature-based algorithm to classify multimedia objects as bearing hidden data or not. Our objective is not to extract the hidden messages or to identify the existence of particular information (as it is in watermarking applications), but only to determine whether a multimedia object was modified by information hiding techniques. Once classified, the suspicious objects can then be inspected in detail by any particular data embedding/retrieving algorithms. This pre-process would particularly contribute to save time in active steganalysis.

As is well known, steganography and watermarking constitute two main applications of information hiding techniques. Though both applications share many common principles in data embedding/extraction schemes, they differ in some criteria, such as robustness, embedding capacity, requirement of original messages, etc. In certain scenarios, content owners might need to determine the existence of hidden watermark in a multimedia object, when the authentication program fails to extract or match the targeted watermarks (due to inversion attack, geometric attacks, de-synchronization attacks etc.). In a possibly negative viewpoint, users may use this steganalytic feature to identify the existence of watermarks in an object. To summarize, steganalysis has promising applications to detect both the steganographic and watermarking schemes.

Our research starts with the analysis and categorization of existing image hiding algorithms. This approach is based on the extension of the fact that hiding information in digital media requires alterations of the signal properties that introduce some form of degradation, no matter how small. These degradations can act as signatures that could be used to reveal the existence of a hidden message. For example, in the context of digital watermarking, the general underlying idea is to create a watermarked signal that is *perceptually identical but statistically different* from the host signal. A decoder uses this statistical difference in order to detect the watermark. However, the very same statistical difference that is created could potentially be exploited to determine if a given image is watermarked or not. The addition of a watermark or message leaves unique artifacts, which can be detected using the various distortion metrics i.e., Image Quality Measures (IQM) [4]. This paper extends the work in [4] and focuses on selecting the content independent features as potential evidences in revealing the presence of hidden messages. We intend to prove that this removal of content dependency enhances the sensitivity of the steganalyzer.

To blindly classify hiding status of an image, we propose an algorithm in which a set of image distortion metrics are defined and utilized to determine the existence of covert channels in the spatial or transformation domain or not. A systematic image database was constructed for algorithm evaluation and a genetic-X-means classifier [42] [43] was trained based on these evaluated features.

## 2 Steganography vs. steganalysis race

Conventional approaches to data hiding within images can be categorized into spatial or transform (e.g., DCT, DWT, Ridgelet etc.) domains [5]. Least Significant Bit (LSB) addition [6],[7],[8] or substitution [10], [11] method is the most popular hiding technique. These techniques operate on the principle of tuning the parameters (e.g., the payload or disturbance) so that the difference between the cover signal and the stego signal is little and imperceptible to the human eyes. Yet, computer statistical analysis is still promising to detect such a distinction that human beings are difficult to perceive. Some tools, such as StegoDos, S-Tools, and

EzStego, provide spatial-domain-based steganographic techniques [2], [5].

There were some spatial-domain steganalytic algorithms [12], [13], [14], [15], [4], [16] developed to be against the above steganographic schemes. Fridrich *et al.* [14] proposed a steganalysis technique based on the fact that bit planes in typical images are more or less correlated so that the LSB plane can be estimated from the other seven ones. This estimation becomes less reliable as the content of the LSB bit plane is further randomized. Kong *et al.* [16] proposed to evaluate the image complexity, following a statistical filter, to determine the existence of secret messages or not, based on the phenomenon that randomization of the LSB bit plane content becomes heavier after information hiding. Sanjay Kumar et al., in [9] discuss an active steganalysis where the estimation about the hidden message length is made. The proposed algorithm reduces the initial-bias, and estimates the LSB embedding message ratios by constructing equations with the statistics of difference image histogram.

Chandramouli *et al.* [12], [13] had ever assumed a Gaussian variation model for LSB disturbances, proposed a maximize *a posterior* (MAP) detector, and analyzed the maximum embedding capacity under which a steganalyst cannot detect the presence of hidden data with a desired probability. Unfortunately, neither detailed implementation of this MAP detector was given nor realistic experiments were reported in their work. Besides, their analysis was restricted to the Gaussian modelling of embedding disturbances. Gokhan Gul et al., in [46] briefly describe PQ and propose singular value decomposition (SVD)-based features for the steganalysis of JPEG-based PQ data hiding in images. They show that JPEG-based PQ data hiding distorts linear dependencies of rows/columns of pixel values, and proposed features can be exploited within a simple classifier for the steganalysis of PQ. Andrew A. Ker [18] proposes more accurate attacks on LSB embedding through a weighted stego image detector for finding the sequential image replacement.

Hiding can also be performed in the transform domain, e.g., DCT [19], [20], [21], [22], [23], [24], [25] or DWT domain [23], [26]. Regardless of which domain, "significant" transform coefficients are often selected to mix with secret/perturbing signal in a way such that information hiding or watermarking is transparent to human eyes. For instance, Cheng *et al.* [24] proposed an additive approach to hiding secret information in the DCT and DWT domains. Wu *et al.* [25] proposed a two-level data embedding scheme, in principle of additive spread spectrum and spectrum partition, for applications in copy control, access control, robust annotation, and content-based authentication. There exist some tools, such as J-Steg and Outguess, providing this category of steganographic techniques [5], [15].

Some steganalytic methods [14], [27], [28] were proposed in the DCT domain. Manikopoulos *et al.* [27] applied the differences in the coefficients of the block DCT transforms of the original as features to the detection of block DCT-based steganography in gray-scale images. The model utilizes statistical pre-processing, over an observation region of each image that generates feature vectors over the regions. These vectors are then fed into a simple neural network classifier. Fridrich *et al.* [14] described that a modified image block will most likely become saturated (i.e., at least one pixel with the gray value 0 or 255) in a JPEG-format stego-image after information hiding. If no saturated blocks can be found, there will be no secret messages therein. Otherwise, a spatial-domain steganalytic method [14] mentioned earlier can be used to analyze these saturated blocks. In [28], the author modelled the common steganographic schemes as a linear transform between the cover and stego images, which can be estimated after at least two copies of a stego image were obtained. This is similar to a blind source separation problem that can be solved by using the independent component analysis (ICA) [29] technique. In [30], a steganalytic scheme was devised to deal with information hiding schemes mixing a secret and a cover signal in an addition rule. The phenomenon, that the center of mass of the histogram characteristic function in a stego image moves left or remains the same to that of the cover image, was observed and exploited to distinguish stego images from plain ones.

It is noticed that most of the steganalytic schemes were designed either in specific operating domain, or even for particular steganographic algorithm. Building a universal steganalytic system is, up to now, a challenging exercise.

In [31], Wen et al. has modelled a universal steganalyzer that operates to distinguish stego images from clean images using two features only namely gradient energy and statistical variance of the Laplacian parameter. The system lacks the ability to strongly attack a wavelet based stego systems. But that can be solved by using a feature that is more sensitive to such embedding strategy. In [44] Der et al. proposes an universal steganalysis scheme that focuses on the differences of statistical features formed by embedding algorithms and applies a support vector machine to distinguish the stego-image from suspicious images. Even though many steganalytic systems have been developed, each system only identifies a subset of the available embedding methods and with varying degrees of accuracy. Benjamin in [45] applies Bayesian model averaging to fuse multiple steganalysis systems and identify the embedding used to create a stego JPEG image.

There are several fundamental questions one may ask:

Which features contribute more to the discriminating power of the universal steganalyzer?

Until what point does steganalysis performance improve with the number of features used? These questions are all related to a crucial ingredient of any blind steganalyzer.

Avcibas *et al.* [4] proposed a concept that any image will incur quality degradation after smoothing or low-pass filtering and this degradation (reacting on image quality) depends on the type of the test image, especially

in categories of with or without embedded information. That is, by observing quality difference between a test image and its smoothed version, it is possible to discriminate images with and without hidden messages. They hence utilized a regression analysis with several quality measuring operators for steganalysis. They have analyzed 26 image quality metrics for the purpose of discrimination. All features are not equally valuable to the learning system. Furthermore, using too many features is undesirable in terms of classification performance due to the curse of dimensionality [29]: one cannot reliably learn the statistics of too many features given a limited training set. Hence, we need to evaluate the features' usefulness and select the most relevant ones.

However it was discovered that removing the inherent content dependency in distortion measures as calculated in [4] is beneficial. So we propose a novel method to remove this content dependency from distortion measurements. These content-independent measurements are then used to build a classifier to differentiate cover-signals and stego-signals. The experimental results justify how the proposed technique enhances the discriminatory power of the features used in the classifier.

# 3 Effect of removing content dependence features

This paper quantifies steganalysis task in the information-theoretic prescription context of data hiding i.e., hiding in independent and identically distributed Gaussian host samples [3]. It is quite common to choose the embedding signal i.e., message to be conveyed as a zero mean, white Gaussian process with finite variance. It is known from information theory that a Gaussian signal is the best choice for a Gaussian channel. Since most image steganography methods conveniently assume the image pixel distribution and common transform coefficient distribution to be Gaussian, the choice of secret message as Gaussian is justified.

The prospects of certain image quality metrics in envisaging the presence of watermarking and steganographic signals within an image is described in [4]. The presence of the steganographic artifact can then be put into evidence by recovering the original cover signal, or alternatively, by de-noising the suspected stego-signal. The steganalyzer can directly apply a statistical test on the denoising residual, $x - \hat{x}$, where $\hat{x}$ is the estimated original signal. This residual must also correspond to the artifact due to embedding of a hidden message. Notice that, even if the test signal does not contain any hidden message, the de-noising step will still yield an output, whose statistics can be expected, however, to be different from those of a true embedding. There subsists a motive to utilize more than one distortion measure, in order to investigate different quality aspects of the signal, which could be brunt during data hiding manipulations. In pursuing such a task, there is often the risk that the variability in the signal content itself surpasses the detector from the alterations. Thus, it is required that, whatever features are selected, the detector responds only to *the induced distortions,* which is Gaussian distributed [3], during data hiding and not be confused by the statistics of the signal content. Moreover, the original signal apparently will not be available during the testing stage. Therefore, some reference signal must be created that is common to both the training and testing stages.

In [4], a denoised version of the given signal is used as the reference. Anyway, this self-referencing, which is creating a reference signal via its own denoised version, is obviously a content-dependent scheme. The classifier performance can be inferior as it responds to both the signal content based statistics and to the distortions stimulated by data embedding operation. To eliminate this content dependency, it is recommended to use a single reference signal that is common to all signals to be tested. Thus, a content independent reference signal and its altered versions according to the type of data embedding are employed.

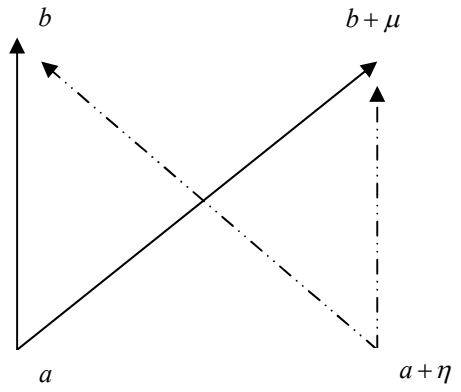


Figure 1: Signal vectors: original signal $a$, its embed-ded version $a+\eta$, reference signal $b$ and its embed-ded version $b+\mu$.

Let $a$ denote a test signal and $a+\eta$ be its stego version, and similarly, let $b$ and $b+\mu$ indicate the reference signal and its stego version. Besides, let us consider a generic distortion functional $Distortion(x, y)$ between the signals $x$ and $y$. For example, for the mean-square distortion, one simply has $Distortion(x, y) \triangleq E[(x-y)^2]$, with $E$ being the expectation operator. The detector operates on the basis of the statistical differences of the distortions. This implicitly ensues two assumptions. First, data embedding leads to additive distortion, that is, the altered signals can be represented as $a+\eta$ and $b+\mu$. Second, the additive distortions of the test and reference signals should not be mutually orthogonal, that is, $E\{\eta, \mu\} \neq 0$. This assumption was indirectly justified by analysis of variance (ANOVA) [4] and the test results given in the experimental results section.

It is to be shown that self-referencing, as employed in [4], causes content-dependent distortion. Let $\Re$ be the specific operation by which the reference signal is generated; for example, in [4], denoising operation has

been used $b = \Re(a) = denoise(a)$. The outcomes of this operation are given by $a \xrightarrow{\Re} \Re(a)$ and $a+\eta \xrightarrow{\Re} \Re(a+\eta)$, respectively, for (1) gnal and its stego version. To illustrate the point, for the case of the mean-square distortion, one obtains

$$Distortion(a+\eta,\Re(a+\eta)) - Distortion(a,\Re(a)) = E[\Re(a+\eta)^2 + 2a\eta + \eta^2 - 2(a+\eta)\Re(a+\eta) + 2a\Re(a) - \Re(a)^2]$$

which is content dependent, because the signal terms $a$ and $\Re(a)$ survive in the difference of distortion functions. The above difference should be some function of only the distortion term and should not contain $a$ or any of signal derived from it, to ensure content independence and to be a real indicator of data embedding effects.

We propose an alternative way and suggest to consider an unique signal $b$ as a reference signal. Then the distortion metrics can be measured between $a$ and $a+\eta$, using $b$ and $b+\mu$ as reference signals. The relationship of these signals and the distortion *vis-à-vis* the reference signals $b$ and $b+\mu$ illustrated in Fig. 1. In this figure, the length of the vector $\overrightarrow{ab}$ is simply equal to $Distortion(a,b)$. The distance between the tips of the vectors $\overrightarrow{ab}$ and $\vec{a}(b+\mu)$ is $d = Distortion(a,b) - Distortion(a,b+\mu)$ and simi-larly $d' = Distortion(a+\eta,b) - Distortion(a+\eta,b+\mu)$, denotes the distance between the tips of the dashed pair of vectors.

For the case of mean-square distortion it follows that

$$d = E[(a-b)^2 - (a-b)^2 + 2(a-b)\mu - \mu^2] = E[2(a-b)\mu - \mu^2] \quad (2)$$

$$d' = E[(a+\eta-b)^2 - (a+\eta-b)^2 + 2\mu(a+\eta-b) - \mu^2] = E(2\mu(a-b) + 2\mu + \eta - \mu^2] \quad (3)$$

To remove the content dependency it is enough that we calculate the difference between $d$ and $d'$.

$$D = 2E[\mu\eta] \quad (4)$$

The same effect of eliminating content dependency can be shown with another distortion metric, the correlation coefficient given by $Distortion(x,y) = E[(xy)]$.

Here $d = E[ab] - E[a(b+\mu)] = -E[a\mu]$ and

$$d' = E[(a+\eta)b] - E[(a+\eta)(b+\mu)] = -E(a\mu] - E[\eta\mu] \quad (5)$$

The removal of content dependency can be shown as the difference between $d$ and $d'$ like $D = d' - d = -E[\eta\mu]$. (6)

# 4 Design of the steganalyzer

This paper mainly concentrates on designing a blind steganalyzer that can distinguish between a clean image and an adulterated image, using an appropriate set of content independent IQMs. Objective image quality measures are based on image features, a functional of which, should correlate well with subjective judgment, that is, the degree of (dis)satisfaction of an observer [32]. Objective quality measures have been utilized in coding artifact evaluation, performance prediction of vision algorithms, quality loss due to sensor inadequacy etc. [33]. In [4] they have extensively studied the use of image quality measures specifically as a steganalysis tool, that is, as features in detecting watermarks or hidden messages.

## 4.1 Content Independent Image Quality Metrics (CIIQMs) as features

A good IQM should be accurate, consistent and monotonic in predicting quality. In the context of steganalysis, *prediction accuracy* can be interpreted as the ability of the measure to detect the presence of hidden message with minimum error on average. Similarly, *prediction monotonicity* signifies that IQM scores should ideally be monotonic in their relationship to the embedded message size or watermark strength. Finally, *prediction consistency* relates to the quality measure's ability to provide consistently accurate predictions for a large set of watermarking or steganography techniques and image types. This implies that the spread of quality scores due to factors of image variety, active warden or passive warden steganography methods should not eclipse the score differences arising from message embedding artifacts. In order to understand how these metrics measure up to the above desiderata [4] resorted to analysis of variance (ANOVA) techniques. The ranking of the goodness of the metrics was done according to the F-scores in the ANOVA tests to identify the ones that responded most consistently and strongly. In the final analysis a list of IQMs is obtained that are sensitive specifically to steganography effects, that is, those measures for which the variability in score data can be explained better because of some treatment rather then as random variations due to the image set.

The stego-detector we develop is based on analysis of a number of *relevant but content independent* IQMs. The idea behind detection of watermark or hidden message presence is to obtain a consistent distance metric for images containing a watermark or hidden message *vis-à-vis* those without, *with respect to a common reference*. The reference processing should possibly include a general signal common to both testing and training. Our approach differs from [4] in using a random signal as the common reference signal rather than using a denoised signal.

The quality metrics exploited in [4] are categorized into six groups according to the type of information they use. The categories used are:

**Pixel Difference-based Measures:** Mean square error, Mean absolute error, Modified infinity norm, L*a*b perceptual error, Neighborhood error and Multi-resolution error.

**Correlation-based Measures***:* Measures based on correlation of pixels, or of the vector angular directions like Normalized cross correlation, Image fidelity, Czenakowski correlation, Mean angle-magnitude similarity and Mean angle similarity.

**Edge-based measures:** Measures based on the displacement of edge positions or their consistency across resolution levels like Pratt edge measure and Edge stability measure.

**Spectral distance-based Measures:** Measures based on the Fourier magnitude and/or phase spectral discrepancy on a block basis like Spectral phase error, Spectral phase-magnitude error, Block spectral magnitude error, Block spectral phase error and Block-spectral phase-magnitude error.

**Proposed Algorithm:** *CIIQM Based Steganalyzer*

***Phase:*** *Learning*

***Input:*** *A database of images*

***Output:*** *A knowledge base capable of discriminating between a clean and a stego image*

1. **Image data base construction**: Prepare an image data base containing clean images and stego images generated out of difference embedding schemes.
2. **Removal of content dependency:** A single random reference signal that is common to all the signals is selected for evaluating IQMs
3. **IQM evaluation:** The various IQMs mentioned in Section 5 like Pixel Difference-based Measures, Correlation-based Measures, Edge-based measures, Spectral distance-based Measures and Context-based Measures are evaluated, between the test signal and the common reference signal.
4. **Genetic Algorithm based feature selection**: The content independent features that are sensitive to data embedding operation are selected based on genetic search strategy. Genetic-X-means algorithm is described in the listing 2.
5. **Data set formation**: A data set is formed out of the selected features.
6. **Training:** The steganalyzer is subjected to learning by applying the X-means algorithm over this data set and a knowledge base is constructed.
7. **System Ready**: The steganalyzer system is now ready for universal blind steganalysis.

***Phase:*** *Detecting*

***Input:*** *A test signal which is to be categorized as a clean or stego bearing image.*

***Output:*** *Categorization of the signal as clean or stego-bearing*

1. **Network Daemon:** It monitors traffic and channelises the multimedia data to the IIU.
2. **Image Identification Unit:** This is used for identification of the image data files. It is achieved by observing the header information of each and every incoming data packet. Various image files being identified by this component are .BMP, .GIF, .TIFF, .PNG etc.
3. **Common reference signal selector:** The same signal chosen in the learning phase is chosen to evaluate the CIIQMs.
4. **CIIQM Evaluator:** The content independent IQMs selected in Step 4 of the learning phase are evaluated against the same common reference signal chosen in the learning phase.
5. **Genetic-X-means Clustering Engine**: The derived feature vector is given to the X-means classifier engine for diagnosis. Based on the knowledge constructed in learning phase this component decides whether the document is adulterated or untouched and identifies the specific steganographic technique used. The algorithm is given in the Listing 2.
6. **Actioner:** Actioner's role is to take necessary actions when a marked document is detected. When an adulterated file is detected exactly, the actioner does one of the following operations. 1. Warn the system administrator 2. Warn the end user 3. Kill the specific application, which executed that image file 4. Prevent the end user from running any further application.5. Extract the hidden information from/in the image file. Case 2, 3, 4 & 5 can be achieved locally at the client workstation.

Listing 1. Framework of the CIIQM based Steganalyzer.

***Algorithm*** **:** *Genetic-X-means algorithm applied to image steganalysis.*

***Input*** **:** *Training set D={(x1,y1),(x2,y2),…:(xn,yn)}, Lower bound=α ,Upper bound=β .*

***Output*** **:** *The clustered model with the maximum BIC (Bayesian Information Criterion) Score, the respective value of K and K centroid parameters.*

1. *Genetic_Feature_Selection()*

Initialize K= α,

2. **Improve Params**: Run *direct k-means* to convergence, with the features selected using genetic search algorithm

**Improve Structure**: Add new centroids where needed by applying SPLITLOOP system as in Gaussian Mixture model identification [43].

For the locally evolved model Mj, evaluate the BIC Score locally evolved, with k=1 and k=2 :

$$BIC(M_j) = i_j(D) - \frac{p_j}{2}\log R$$

where $i_j(D)$ is the log-likelihood of the data according to the j-th model and taken at the maximum-likelihood point, $p_j$ is the number of parameters in $M_j$ and $R = |D|$.

Sustain the model $M_j$ having greater BIC score.

Record the parameters of the model evolved, like: K, BIC score of the entire model, K Centroid values.

8. If K< β, Goto 2.

Listing 2: Pseudo code for Genetic-X-Means algorithm applied to Image Steganalysis.

**Context-based Measures:** Measures based on penalties for various functional of the multidimensional context probability like Rate distortion measure, Hellinger distance, Generalized Matusita distance and Spearman rank correlation.

## 4.2 Choice of genetic-X-means paradigm

According to whether the steganalysis is based on supervised or unsupervised learning, stego-anomaly detection schemes can be classified into two categories: unsupervised stego-anomaly detection and supervised stego-anomaly detection. In supervised strategy, profiles of clean/stego files are established by training using a labelled dataset. Unsupervised detection uses unlabelled to identify anomalies. The main drawback of supervised detection is the need to label the training data, which makes the process error-prone, costly and time consuming. Unsupervised anomaly detection addresses these issues by allowing training based on an unlabelled dataset and thus facilitating online learning and

**Algorithm** **:** *direct K-means()*

**Inputs:**

*I = {i1,i2,…ik} (Stego/Clean instances to be clustered)*

*n (Number of Clusters)*

$\theta$ *(Threshold value as a stopping criterion)*

**Outputs:**

*C = {c1,c2,…,ck} (Cluster centroids)*

*m:* $I \to C$ *(Cluster membership)*

Initialize k prototype $(w_1, w_2 \cdots w_k)$ such that $w_j = i_l$ , $j \in \{1,2,...,k\}$ , $l \in \{1,2,...,n\}$ .

Each cluster $C_j$ is associated with prototype $w_j$

Repeat

For each input vector $i_l$, where $l \in \{1,2,...,n\}$ do

$m(i_j) = \arg\min_{k \in \{1..n\}} distance(i_j, c_k)$

For each cluster $C_j$ ,where $j \in \{1,2,...,k\}$ do

Update the prototype $w_j$ to be the centroid of all samples currently in $C_j$ so that $w_j = \sum_{i_l \in C_j} i_l / |C_j|$ and $m(i_j) = \arg\min_{k \in \{1..n\}} distance(i_j, c_k)$

Compute the error function

$$E = \sum_{j=1}^{K} \sum_{i_l \in C_j} |i_l - w_j|^2$$

Until $E < \theta$ or cluster membership no longer c hanges.

Listing 3: Pseudo code for direct k-means algorithm.

improving detection accuracy. By facilitating online learning, unsupervised approaches provide a higher potential to find new attacks. By removing the need of labelling, unsupervised detection creates a greater potential for accurate detection.

Clustering is the organization of data patterns into groups or clusters based on some measure of similarity. When applying clustering techniques for steganalysis, determining the number of clusters is a difficult issue since the data hiding algorithm is unknown. The general approach and current practice assume that data instances are always divided into two categories: normal clusters and anomalous clusters. However, this assumption need not always true in practice. The number of clusters is not supposed to be determined in advance. When data instances include only normal behavioral data, the assumptions will lead to a high false alert rate and a vice-versa case when data instances include only stego patterns. In order to achieve an efficient and effective detection, we propose in this paper, a new unsupervised stego-anomaly detection framework which consists of a clustering algorithm, named X-means and a CIIQM feature extraction based on genetic search. X-means

**Algorithm :** *Genetic_Feature_Selection()*

**Input :**
*Encoded binary string of length 26 (one bit for each IQM), number of generations, and population size, Cross over probability Pc, Mutation Probability Pm.*

**Output :**
*A set of selected features.*

1. Initialize the population randomly

2. $W1 = 10^4$, $W2 = 0.4$,

3. $N$ = total number of records in the training set
4. For each chromosome in the new population

5. Apply uniform crossover operator to the chromosome with a probability of Pc.6. Apply mutation operator to the chromosome with a probability of Pm.

7. Evaluate Fitness = *W1 * Accuracy + W2 * Zeros*

8. Select the top best 50% chromosomes into new population using Tournament Selection operator.

9. If number of generations is not reached, go to step 4.

Listing 4: Pseudo Code for Feature Selection by Genetic-Search Strategy.

algorithm extends appropriately k-means with some evolutionary steps, integrates the capability of determining automatically the optimal number of clusters for a set of data, and thus addresses the limitation of traditional clustering based intrusion detection approaches.

# 5 Proposed algorithm for CIIQM genetic-X-means image steganalyzer model

The proposed algorithm for genetic-X-means based image steganalysis system is provided here. This can be set up in the network of the corporate sectors. The multimedia traffic (image, video, image, text, HTML pages etc.,) is keenly monitored by the system. Whenever the entry of image documents is sensed, the steganalyzer is triggered. The system consists of two main stages. They are 1. Learning Stage 2. Detecting Stage.

# 6 Experimental topology

In our experiments, the discrimination performance of content independent features is analyzed first. Then the classification performance of our steganalyzer under the prepared test image set is reported. Besides, the impacts of embedding rate and mismatch between the training and test sets (e.g., modified by using different embedding schemes) on the classification rate are also explored.

## 6.1 Preparation of test images and schemes

The design of experiments is important in evaluating our steganalytic algorithm. The key considerations include the following.

1) First, from the point of "generalization", the proposed content independent image features and associated classifier should be capable of identifying the existence of hidden data which are possibly generated by using various kinds of embedding methods, regardless of steganography or watermarking, and regardless of spatial or transform-domain operations.

2) Second, in outlook of "performance", the classifier should, on the one hand, detect hidden data as likely as possible (regardless of how transparent the embedded secret information is), and on the other hand, keep false alarms to as few as possible for plain images.

3) Third, in view of "robustness", the classifier should be capable of differentiating the effect of ordinary image processing operations (such as filtering, enhancement, etc.) from that of data embedding.

On the grounds of the above considerations, six published methods based on two types of principles, LSB embedding and spread spectrum, were chosen for evaluation.

scheme **#1**: Digimarc [34]
scheme **#2**: PGS [35]
scheme **#3**: Cox *et al.*'s [22]

scheme **#4**: S-Tools [36]

scheme **#5**: Steganos [37]
scheme **#6**: JSteg [38]
scheme **#7:** Kim *et al.'s* [39]

They can be further categorized into:

1) steganography (#4, #5, #6) or watermarking (#1,#2,#3) purpose;

2) spatial (#2, #4, #5), or transform (#1, #3, #6) domain operation.

For further testing and to verify the effectiveness of the features selected, we select an extra scheme based on the wavelet domain:

3) scheme **#7**: Kim *et al.*'s method [39].

It is expected that the difference between a cover image and its stego version can be easily detected when more secret messages are embedded. Hence the capacity of the payload of a steganography scheme should be taken into account in evaluating the detection capacity of a steganalytic classifier. To depict this, the embedding rate (ER) characterizing a scheme which is defined as the ratio between the number of embedded bits and the number of pixels in an image, is used.

- Mean square error
- Median block weighted spectral distance
  HVS based L2
- HVS normalized absolute error
- Weighted spectral distance
- Cross correlation.

Table I: content independent distortion measures selected by genetic search

The wavelet-based steganography scheme #7 was used to test our steganalytic scheme, although the proposed features are trained only on the spatial and the DCT domains.

To test the performance of the proposed method, our cover image dataset consists of 200 with a dimension of 256 X 256 8-bit gray-level photographic images, including standard test images such as Lena, Baboon, and also images from [40]. Our cover images contain a wide range of outdoor/indoor and daylight/night scenes, including nature (e.g., landscapes, trees, flowers, and animals), portraits, manmade objects (e.g., ornaments, kitchen tools, architectures, cars, signs, and neon lights), etc. Some of the sample images are shown in Fig. 2. This database is augmented with the stego versions of these images using the above mentioned seven schemes, at various embedding rates. Also a separate image set was generated by applying the image processing techniques like JPEG compression (at several quality factors), low-pass filtering, image sharpening etc. Our generation procedure is aimed at making even contributions to database images from different embedding schemes, from original or stego, and from processed or non-processed versions, so that the evaluation results can be more reliable and fair. Three different ERs are attempted for each scheme in generating the database like (#1) 5% (#2) 10% (#3) 20% of the maximum payload capacity prescribed by the techniques. The entire database contains 200*4*7=5600 (No. of images * No. of varying ER - 3 ER + 1 for clean set* No. of schemes evaluated) images on the whole.

## 6.2 Content independent features selection

Applying the proposed methodology and the algorithm, the content dependency was removed and the six measures as in Table I are selected after removing the content dependency from the signal.

## 6.3 Feature discrimination capability

Before proceeding to evaluate the performance of the classifier, discrimination capability of the proposed features is to be analyzed. The experiment involves breaking of different steganographic or watermarking strategies, which may adapt extremely different techniques for embedding ranging from LSB substitution to embedding inside the wavelet co-efficient.

Hence the feature set formed has to be normalized before feeding into the classifier for training to achieve a uniform semantics to the feature values. A set of normalized feature vectors as per the data smoothing function [41],

$$\breve{f}_i = \frac{f_i - f_i^{\min}}{f_i^{\max} - f_i^{\min}} \ , \qquad (7)$$

are calculated for each seed image to explore relative content independent feature variations after and before it is modified. $\breve{f}_i$, $f_i^{\min}$ and $f_i^{\max}$ represents the i[th] feature vector value, the corresponding feature's minimum and maximum value respectively.

## 6.4 Genetic-X-means classifier

In the sequel, the model is incorporated in Java JGAP [42] and the algorithm described in section [6] is implemented as per the framework proposed. The classifier was trained and evaluated by using 4800 images out of the whole database, excluding those generated by using scheme #7 (employed as the test images to see how the proposed features behave when there is a mis-match between the operation domains). Here, two-thirds (3200) of images were randomly chosen as the training set and the others (1600 images) act as the validation set.

Before evaluation, some performance indices are first defined.

- Positive detection (PD)—classifying the stego images correctly.

- Negative detection (ND)—classifying the non-stego images correctly.
- False positive (FP)—classifying the presence of secret information for non-stego images.
- False negative (FN)—bypassing or ignoring the presence of hidden information in stego images.

The classification and error rates obtained by using different values are listed in Table II. Results show that the average classification rate does not change much (from 79.5% to 86.67%). We are interested in analysing the detectability of proposed features and classifier against embedding schemes of different applications or principles. Table III lists classification and error rates to see differentiation in performances between: 1) six targeted embedding schemes; 2) steganographic or watermarking applications; 3) spatial or DCT operation domain; and 4) types of processed non-stego images. We also analyzed the ND rates for the original, smoothed, sharpened, and JPEG-compressed non-stego images. It is found that our system has a better performance in recognizing the plainness of JPEG-compressed images. The higher ND rate for JPEG-compressed images is beneficial to real applications, since most images will be compressed in the JPEG form.

| | **PD** | | **ND** | | **Classification Rate(PD+ND)/2** | |
|---|---|---|---|---|---|---|
| **Scheme** | **IQM** | **CIIQM** | **IQM** | **CIIQM** | **IQM** | **CIIQM** |
| DigiMarc | 80% | 85.63% | 80% | 84.97% | 80% | 85.30% |
| PGS | 80% | 81.02% | 90% | 92.31% | 85% | 86.67% |
| Cox | 80% | 84.96% | 60% | 72.30% | 70% | 78.63% |
| S-Tools | 90% | 93.31% | 60% | 78.04% | 75% | 85.68% |
| Steganos | 80% | 87.63% | 60% | 75.54% | 70% | 81.59% |
| Jsteg | 70% | 84.97% | 70% | 74.02% | 70% | 79.50% |
| | **FP** | | **FN** | | **Error Rate (FP+FN)/2** | |
| **Scheme** | **IQM** | **CIIQM** | **IQM** | **CIIQM** | **IQM** | **CIIQM** |
| DigiMarc | 20% | 14.37% | 20% | 15.03% | 20% | 14.70% |
| PGS | 10% | 19.98% | 20% | 7.69% | 15% | 13.84% |
| Cox | 20% | 15.04% | 40% | 27.70% | 30% | 21.37% |
| S-Tools | 10% | 6.69% | 40% | 21.96% | 25% | 14.33% |
| Steganos | 20% | 12.37% | 40% | 24.46% | 30% | 18.42% |
| Jsteg | 30% | 15.03% | 30% | 25.98% | 30% | 20.51% |

Table 2: Performance comparison of the classifiers.

| Differentiation categories | | PD rate |
|---|---|---|
| Schemes | #1 | 85.3% |
| | #2 | 86.67% |
| | #3 | 78.63% |
| | #4 | 85.68% |
| | #5 | 81.59% |
| | #6 | 79.5% |
| Applications | Watermarking | 83.53% |
| | Steganography | 82.25% |
| Operation domain | Spatial | 84.64% |
| | DCT | 81.14% |
| | DWT | 86.32% |
| Differentiation categories | | ND rate |
| Type of processed non-stego images | Original | 82.34% |
| | JPEG-compressed | 89.40% |
| | Smoothed | 83.50% |
| | Sharpened | 58.10% |

Table 3: Average pd/nd rates for performance differentiation between different target schemes, different applications, different operation domains, and different types of nonstego images.

As for the detectability between different embedding schemes, we compare scheme #4 to #5 and scheme #1 to #3. Basically, embedding schemes #4 and #6 are similar in some aspects (both are in the spatial-domain, but for different applications), but the pixel change will be less for scheme #4 when embedding "0." Accordingly, we got a higher PD rate for scheme #4 than for scheme #5.

## 6.5 Influence of embedding rate

In this experiment, the images at various payload capacities were selected to see the influence on detectability. The ERs for the six embedding schemes were tried at 5%, 10% and 20% of the maximum hiding capacities in their proposed versions. The experimental results are listed in Table V, which depicts that the average PD rate still remains above 83.38% for 20%, 78.78% for 10% and 74.47% for 5% of maximum payload capacity. The results for steganographic schemes are more promising than for the watermarking schemes, as the steganographic schemes carry more hidden data than those of watermarking schemes, which makes the measured features more distinguishable for detection. The results reveal that clearly, our proposed content independent features and genetic-X-means classifier still yield reasonable results for stego images of less ER.

## 6.6 Detection with mismatch between the training and test sets

Here we evaluate the performance when images modified by using different kinds of hiding schemes are employed for training and testing. Denote the training set and the test set as $S_L$ and $S_T$ respectively.

First, we created $S_L^1$ by including stego images generated by using the steganographic schemes #1 and some processed plain images. On the other hand, $S_T^1$ is constituted of stego images produced by using the watermarking schemes #2, #3 and other processed plain images. Essentially, $S_L^1$ and $S_T^1$ were made disjoint and consist of 400 and 600 images, respectively. We also evaluate the detection performances in presence of other mismatches In the second case, we interchanged the roles of $S_L^1$ and $S_T^1$ to form another two sets, $S_L^2$ and $S_T^2$, i.e., $S_L^2 \leftarrow S_T^1$ and $S_T^2 \leftarrow S_L^1$. Similarly $S_L^3$ includes stego images created using schemes #4 and #5. $S_T^3$ constitutes the stego images processed by using scheme #6. $S_L^4$ and $S_T^4$ represent the reversed role of $S_L^4$ and $S_T^4$ sets. Another set $S_L^5$ and $S_T^5$ contains the stego images created by employing the schemes #1,#2,#4,#5 and tested on schemes #3 and #6 respectively. Their interchanged sets are $S_L^6$ and $S_T^6$.The experimental results are listed in Table IV, which reveals that the average classification rate is 83.35%. Noticeable is that the PD rate for $S_T^2$, $S_T^4$ and $S_T^6$ is much higher than that for $S_T^1$, $S_T^3$ and $S_T^5$. The reason is that the steganographic schemes that embed in the spatial domain reveal more statistical evidence than the ones which hide in the transform

| Data Group | PD | ND | Classification rate | FP | FN | Error rate |
|---|---|---|---|---|---|---|
| $S_L^1, S_T^1$ | 70.71% | 86.69% | 78.7% | 29.29% | 13.31% | 21.3% |
| $S_L^2, S_T^2$ | 75.74% | 87.08% | 81.41% | 24.26% | 12.92% | 18.59% |
| $S_L^3, S_T^3$ | 79.28% | 88.07% | 83.68% | 20.72% | 11.93% | 16.33% |
| $S_L^4, S_T^4$ | 92.13% | 83.37% | 87.75% | 7.87% | 16.63% | 12.25% |
| $S_L^5, S_T^5$ | 73.74% | 89.69% | 81.72% | 26.26% | 10.31% | 18.29% |
| $S_L^6, S_T^6$ | 91.25% | 82.48% | 86.87% | 8.75% | 17.52% | 13.14% |

Table 4: Classification rates obtained when characteristics of the training and test sets mismatch to each other.

| Schemes | Classification rate | | | Error rate | | |
|---|---|---|---|---|---|---|
| | **5% of maximum payload** | **10% of maximum payload** | **20% of maximum payload** | **5% of maximum payload** | **10% of maximum payload** | **20% of maximum payload** |
| #1 | 80.40% | 83.20% | 85.30% | 19.60% | 16.80% | 14.70% |
| #2 | 79.80% | 83.10% | 86.67% | 20.20% | 16.90% | 13.33% |
| #3 | 70.11% | 73.50% | 78.63% | 29.89% | 26.50% | 21.37% |
| #4 | 76.30% | 79.22% | 85.68% | 23.70% | 20.78% | 14.32% |
| #5 | 71.20% | 78.33% | 81.59% | 28.80% | 21.67% | 18.41% |
| #6 | 69.80% | 74.55% | 79.50% | 30.20% | 25.45% | 20.50% |
| #7 | 73.66% | 79.54% | 86.32% | 26.34% | 20.46% | 13.68% |

Table 5: Classification and error rates for test sets at various embedding rate.

domain. This makes the measured content independent features more distinguishable for detection. After examining Table IV, it was found that $S_T^2$ , $S_T^4$ and $S_T^6$ will gain a high PD rate than $S_T^1$ , $S_T^3$ and $S_T^5$ . Hence, we have a conjecture that characteristics (e.g., ER, type of applications, or operating domain) of the test stego image may play an important role on PD rates, but mismatch between the training and test sets might not be so significant.

## 6.7 Application on a completely new steganography scheme

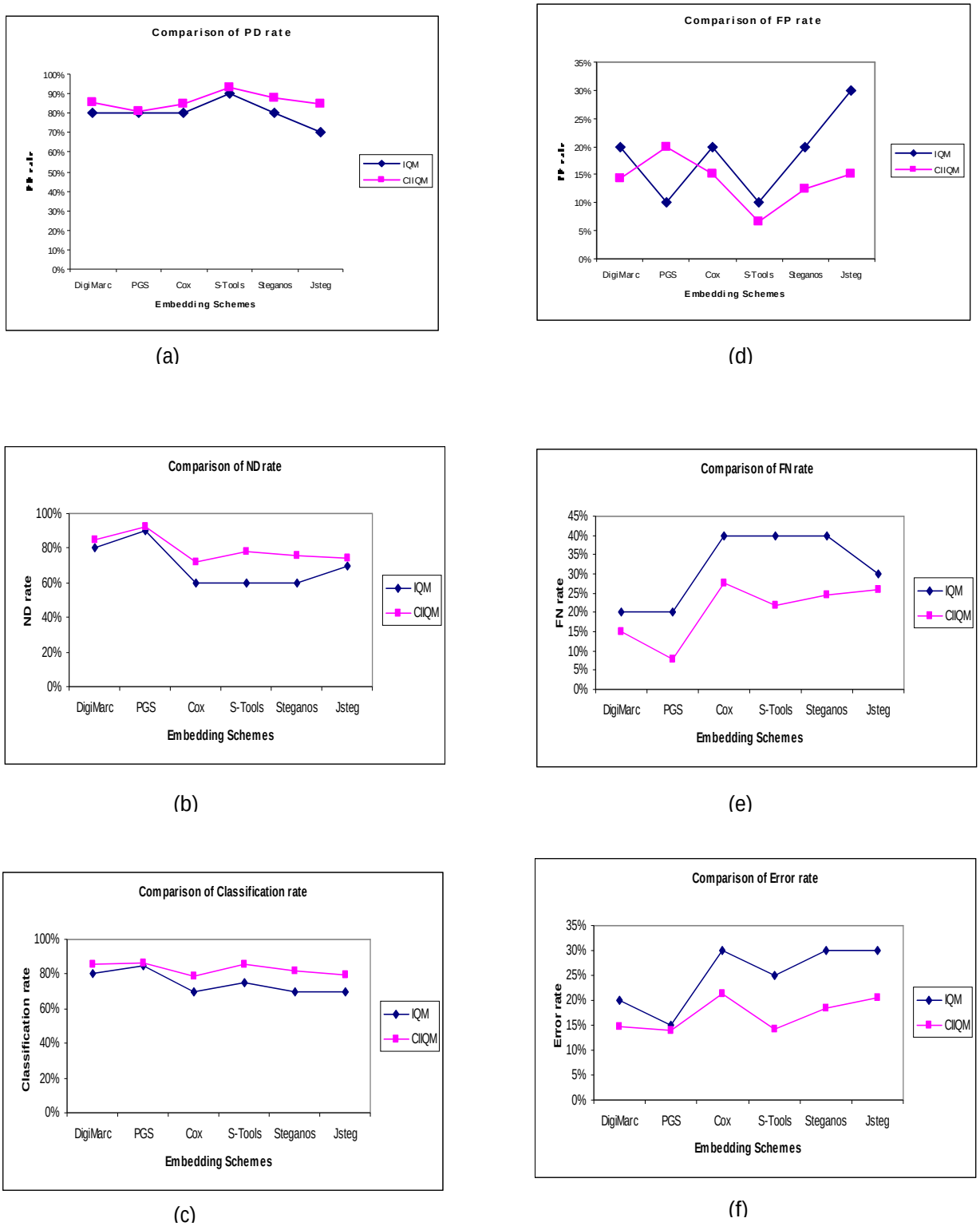


Figure 2: Performance comparison curves depicting a) Positive detection rate b) Negative detection rate c) Classification rate d) False positive rate e) False negative rate f) Error rate

In order to show that the system is dynamic i.e., adaptable to detect any new steganographic technique, the system was tested on scheme #7, which is based on the wavelet-domain techniques.. It was found that the PD rate against scheme #7 is 86.32% as given in Table V. This proves that the identified content independent IQMs are sensitive to detect even any new stego systems. To accommodate the identification of more hiding schemes, other kinds of image features should be explored further.

# 7 Discussion and conclusion

Recently, information hiding techniques find its applications in several fields, e.g., watermarking, copyright protection, steganography, fingerprinting, digital rights management (DRM), etc. At one end there are much research works focusing on addressing the various edges of data embedding techniques like enhancing the transparency, robustness and capacity. On the other end it is, however, interesting to detect the existence of hidden data resulting from any kind of

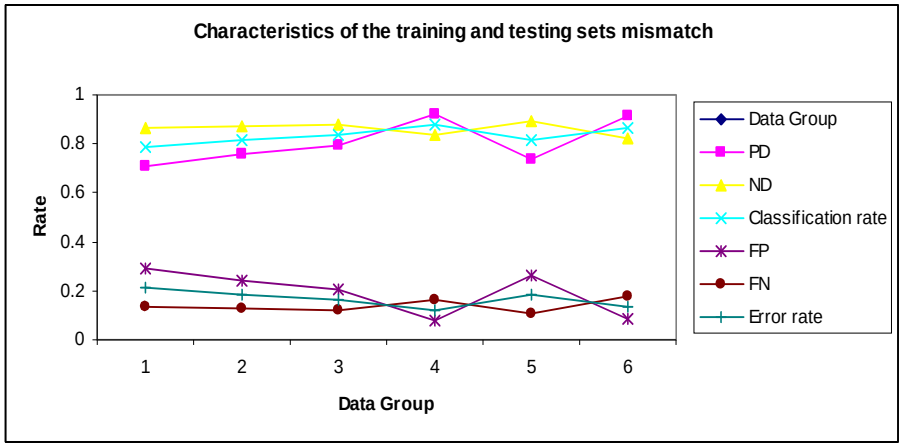


Figure. 3. Detection with Mismatch between the Training and Test Sets.

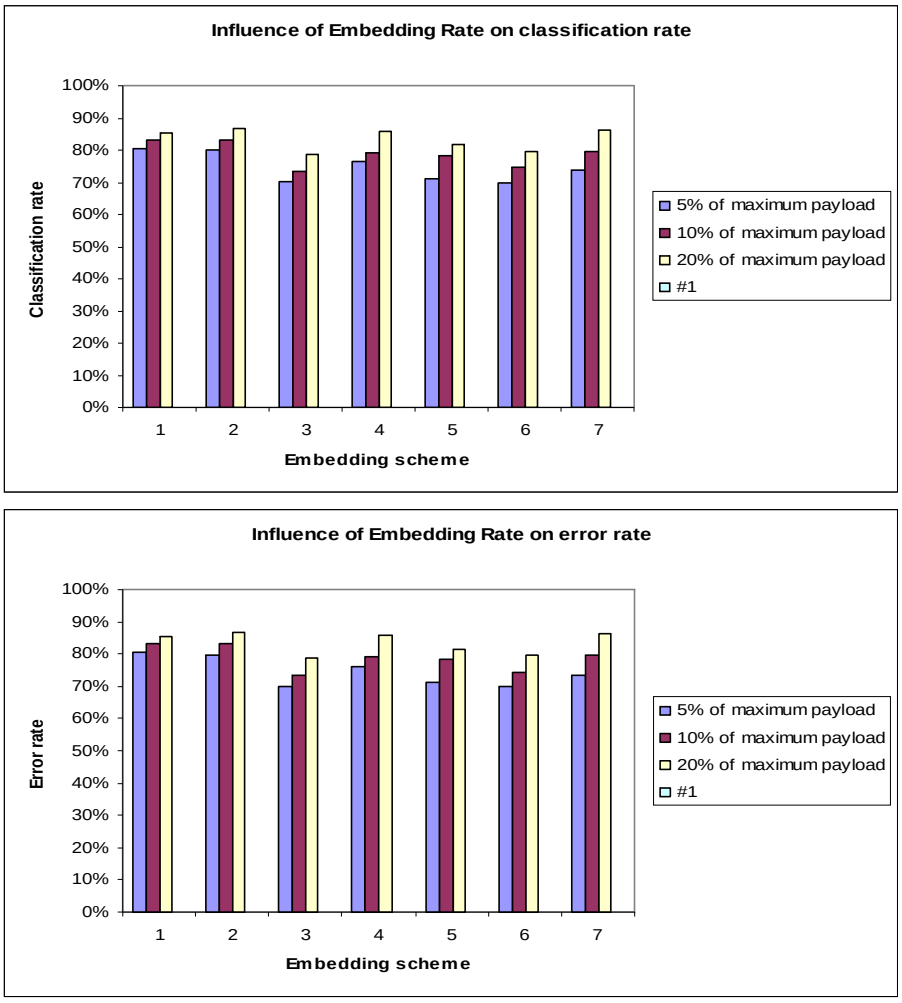


Figure 4. Influence of Embedding rate on performance of the steganalyzer.

embedding scheme, known as the "steganalysis." We have presented a rationale for a content independent distortion metrics based model for blind image steganalysis i.e., without knowledge of the kind of steganographic schemes and shown evidence using systematic experimental results. In our experiments, a database composed of processed plain images and stego images generated by using seven embedding schemes was utilized to evaluate the performance of our proposed features and classifier. Removal of content dependency from the measurements enhanced the classifier's discriminatory power and proved to be useful, especially for steganographic data embedding, where the incurred distortions are much less pronounced than in watermarking.

Table VI summarizes and compares characteristics of our proposed method with those of several other previous works in literature. In the table, not-reported (NRP) represents null information provided by the original work. For clarity, several key points are collected as follows.

1. It seems that the classification performance is necessarily proportional to the removal of content dependency from the features, heavily dependent on the ER and number of embedding schemes under tests.
2. Similar to [31], [36], and [43], our system is operated blindly and not restricted to the detection of a particular steganographic scheme (such as LSB or spread spectrum).
3. A nonlinear classifier that is easy to adapt to non-separable classes is adopted in both [33] and our system. However, the system introduced in [33] was only dedicated to the detection of spread spectrum scheme and few test images were used.
4. Our training and test database collects a larger number of stego and non-stego image samples, that were generated by using different steganographic schemes (seven kinds), different embedding rates (5% –20% maximum payload), and different image processing (low-pass filtering, sharpening, JPEG compression). This diversity makes our system more approximate to real applications.
5. The average classification rate (83%, including the PD and ND rates) for our proposed system is superior to [31] in blind steganalysis research.

To make our system more practical, future work could include the following.

a. Fitting the proposed system to classify compressed images or videos.
b. Identifying the type of steganographic algorithm utilized to generate the stego image and locating the image regions exploited to hide secret messages (active steganalysis). After these, we may be able to locate, retrieve, and analyze the embedded messages to infer the conveyed information.
c. Improving the performance as well as the scalability of the blind steganalyzer using appropriate fusion techniques

# 8 Acknowledgement

This work was supported by grants from National Technical Research Organization of Government of India, as a part of "Smart and Secure Environment". The authors sincerely thank the Management, Principal and Head of the Department of Information Technology of Thiagarajar College of Engineering, Madurai, India, for their support and encouragement. Authors would like to thank the anonymous reviewers for the constructive comments which helped to improve the clarity and presentation of the paper.

| **Steganalytic Systems** | **[13]** | **[14]** | **[27]** | **[16]** | **[4]** | **[30]** | **[31]** | **Proposed** |
|---|---|---|---|---|---|---|---|---|
| Number of features | 1 | 4 | 164 | 2 | 10 | 1 | 2 | 5 |
| Domains of Feature Extraction | Spatial | Spatial DCT | DCT | Spatial | Spatial DFT | DFT | Spatial DCT | Spatial DCT DWT |
| Training/Classifier | Yes/Linear | Yes/Linear | Yes/Neural | Yes/Linear | Yes/Linear | Yes/Linear | Yes/Neural | Yes/Genetic-X-means |
| Targeted embedding scheme | LSB | Arbitrary | Spread spectrum | LSB | Arbitrary | Arbitrary | Arbitrary | Arbitrary |
| Number of test schemes | 1 | 6 | 1 | 1 | 6 | 3 | 6 | 7 |
| Payload of stego images | 0.65 bpp | >0.05bpp | 0.016 bpp | >0.05 bpp | >0.01 bpp | 1 bpp | 0.01-2.66 bpp | >0.01 bpp |
| Size of training database | NRP | 331 | 28 | NRP | 12 | 20 | 1716 | 3200 |
| Number of test images | NRP | NRP | 14 | 80 | 10 | 4 | 572 | 1600 |
| Average PD rate | NRP | NRP | 0 | 97% | 72.08% | 96.2% | 80.28% | 86.25% |
| Average ND rate | NRP | NRP | 0 | NRP | NRP | 94.8% | 79.56% | 79.53% |
| Side information constraint for classifier | No | No | Average PDF of selected plain images. | No | No | No | No | No |

Table 6: Summarization of previous works and our proposed system. *NRP-Not Reported

# 9 References


[1] G. J. Simmons (1984). The prisons' problem and the subliminal channel. *Proc. Advances in Cryptology (CRYPTO'83)*, pp. 51–67.

[2] F. A. P. Petitcolas, R. J. Anderson, and M. G. Kuhn (1999). Information hiding—A survey. *Proc. IEEE*, vol. 87, no. 7, pp. 1062–1078.

[3] R. Chandramouli (2002). A mathematical approach to steganalysis. *Proc. SPIE*, vol. 4675, pp. 14–25.

[4] I. Avcibas, N. Memon, and B. Sankur (2003). Steganalysis using image quality metrics. *IEEE Trans. Image Process.*, vol. 12, no. 2, pp. 221–229.

[5] S. Katzenbeisser and F. A. P. Petitcolas.(2000). *Information Hiding Techniques for Steganography and Digital Watermarking*. Norwood, MA, Artech House.

[6] W. Bender, D. Gruhl, N. Morimot, and A. Lu (1996). Techniques for data hiding. *IBM Syst. J.*, vol. 35, no. 3/4, pp. 313–336.

[7] N. Nikolaidis and I. Pitas (1998). Robust image watermarking in the spatial domain. *Signal Process.*, vol. 66, pp. 385–403.

[8] L. M. Marvel, C. G. Boncelet Jr., and C. T. Retter (1999). Spread spectrum image steganography. *IEEE Transactionsoin Image Processing*, vol. 8, no. 8, pp. 1075–1083.

[9] Sanjay Kumar Jena, G.V.V. Krishna (2007). Blind Steganalysis: Estimation of Hidden Message Length. *International Journal of Computers, Communications & Control,* vol. II 2007.

[10] T.-S. Chen, C.-C. Chang, and M.-S. Hwang (1998). A virtual image cryptosystem based upon vector quantization. *IEEE Transactions on Image Processing*, vol. 7, no. 10, pp. 1485–1488.

[11] Y. K. Lee and L. H. Chen (2000). High capacity image steganographic model. *Proc. Inst. Elect. Eng., Vis. Image Signal Processing*, vol. 147, no. 3, pp. 288–294.

[12] R. Chandramouli and N. Memon (2000). A distribution detection framework for watermark analysis, *Proc. ACM Multimedia*, pp. 123–126.

[13] R. Chandramouli and N. Memon (2001). Analysis of LSB based image steganography techniques. *Proc. Int. Conf. Image Processing*, pp. 1019–1022.

[14] J. Fridrich and M. Goljan (2002). Practical steganalysis of digital images-state of the art. *Proc. SPIE*, vol. 4675, pp. 1–13.

[15] S. Voloshynoskiy, A. Herrigel, Y. Rytsar, and T. Pun (2002). StegoWall: Blind statistical detection of hidden data. *Proc. SPIE*, vol. 4675, pp. 57–68.

[16] X. Kong, T. Zhang, X. You, and D. Yang (2002). A new steganalysis approach based on both complexity estimate and statistical filter.*Proc. IEEE Pacific-Rim Conf. on Multimedia*, vol. LNCS 2532, pp. 434–441.

[17] Gökhan Gül, Ahmet Emir Dirik, and Ismail Avcıbas (2007). Steganalytic Features for JPEG Compression-Based Perturbed Quantization. *IEEE Signal Processing Letters*, vol. 14, No. 3, pp. 205-208.

[18] Andrew D. Ker. (2007). A Weighted Stego Image Detector for Sequential LSB Replacement. Proc. 2007 International Workshop on Data Hiding for Information and Multimedia Security attached to IAS 07. IEEE Computer Society Press.

[19] W.-N. Lie, G.-S. Lin, and C.-L. Wu (2000). Robust image watermarking on the DCT domain. *Proc. IEEE Int. Symp. Circuits and Systems*, pp. I228–I231.

[20] J. Huang and Y. Q. Shi (1998). Adaptive image watermarking scheme based on visual masking. *Electron. Lett.*, vol. 34, no. 8, pp. 748–750.

[21] T. Ogihara, D. Nakamura, and N. Yokoya (1996). Data embedding into pictorial with less distortion using discrete cosine transform. *Proc. ICPR'96*, pp. 675–679.

[22] I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoon (1997). Secure spread spectrum watermarking for multimedia. *IEEE Transactions on Image Processing*, vol. 6, no. 12, pp. 1673–1687.

[23] C. I. Podilchuk and Z. Wenjun (1998). Image-adaptive watermarking using visual models. *IEEE J. Select. Areas Commun.*, vol. 16, no. 4, pp. 525–539.

[24] Q. Cheng and T. S. Huang (2001). An additive approach to transform-domain information hiding and optimum detection structure. *IEEE Transactions on Multimedia*, vol. 3, no. 3, pp. 273–284.

[25] F. P. Gonzalez, F. Balado, and J. R. H. Martin (2003). Performance analysis of existing and new methods for data hiding with known-host information in additive channels. *IEEE Transactions on Signal Process.*, vol. 51, no. 4, pp. 960–980.

[26] Y.-S. Kim, O.-H. Kwon, and R.-H. Park (1999). Wavelet based watermarking method for digital images using the human visual system. *Electronic Letters.*, vol. 35, no. 6, pp. 466–468.

[27] C. Manikopoulos, Y.-Q. Shi, S. Song, Z. Zhang, Z. Ni, and D. Zou(2002). Detection of block DCT-based steganography in gray-scale images. *Proc. 5th IEEE Workshop on Multimedia Signal Processing*, pp. 355–358.

[28] R. Chandramouli (2002). A mathematical approach to steganalysis. *Pro SPIE*, vol. 4675, pp. 14–25.

[29] R. O. Duda, P. E. Hart, and D. G. Stork (2001). *Pattern Classification*. New York: Wiley-Interscience.

[30] J. J. Harmsen and W. A. Pearlman (2003). Steganalysis of additive noise modelable information hiding. *Proc. SPIE*, pp. 21–24.

[31] Wen-Nung Lie and Guo-Shiang Lin (2005). A Feature-Based Classification Technique for Blind Image Steganalysis. *IEEE Transactions on Multimedia*, vol. 7, no. 6.

[32] S. Daly (1993). The visible differences predictor: An algorithm for the assessment of image fidelity. *Digital Images and Human Vision*, A. B. Watson, Ed. Cambridge, MA: MIT Press, pp. 179–205.

[33] C. E. Halford, K. A. Krapels, R. G. Driggers, and E. E. Burroughs (1999). Developing operational performance metrics using image comparison metrics and the concept of degradation space. *Opt. Eng.*, vol. 38, pp. 836–844.
[34] PictureMarc, Embed Watermark, v 1.00.45, Digimarc Corp.
[35] M. Kutter and F. Jordan. JK-PGS (Pretty Good Signature). [Online]. Available: http://ltswww.epfl.ch/~kutter/watermarking/JK_PGS.html, Last retrieved 10, September 2008.
[36] A. Brown. S-tools version 4.0. [Online]. Available: http://members. tripod.com/steganography/stego/s-tools4.html, Last retrieved 11, September 2008.
[37] Steganos II Security Suite.. [Online]. Available: http://www.steganos.com/english/steganos/download.htm, Last retrieved 12, September 2008.
[38] J. Korejwa. Jsteg shell 2.0. [Online]. Available: http://www.tiac.net/users/korejwa/steg.htm, Last retrieved 13, September 2008.
[39] Y.-S. Kim, O.-H. Kwon, and R.-H. Park, "Wavelet based watermarking method for digital images using the human visual system," *Electron. Lett.*, vol. 35, no. 6, pp. 466–468, 1999.
[40] Images. [Online]. Available: http:// www.cl.cam.ac.uk/~fapp2/watermarking/benchmark/image_database.html. Last retrieved 10, September 2008.
[41] Fang Min A Novel Intrusion Detection Method Based on Combining Ensemble Learning with Induction-Enhanced Particle Swarm Algorithm *IEEE Third International Conference on Natural Computation (ICNC 2007)* .
[42] http://jgap.sourceforge.net/, Last retrieved 20, September 2008.
[43] Dan Pelleg and Andrew Moore (2000). X-means: Extending K-means with Efficient Estimation of the Number of Clusters.*ICML 2000.*
[44] Der-Chyuan Lou, Chih-Lin Lin, and Chiang-Lung Liu (2007). Universal steganalysis scheme using support vector machines. *Optical Engineering*. vol. 46, 117002.
[45] Benjamin Rodriguez, Gilbert Peterson and Kenneth Bauer (2008). Fusion of Steganalysis Systems Using Bayesian Model Averaging. *IFIP International Federation for Information Processing* Springer Verlag, 2008.