A Review on Steganalysis Techniques: From Image Format Point of View

Sherif M. Badr
Computer Science Dept.
Modern Academy, Cairo, Egypt

Gouda I. Salama
Computer Engineering Dept.
MTC, Cairo, Egypt

Gamal M. I. Selim
Computer Engineering Collage
AAST, Cairo, Egypt

Ashgan H. Khalil
Faculty of Computer Science
Modern University, Cairo, Egypt

ABSTRACT
Steganalysis is the art and science of detecting the embedded message in a multimedia document, this document may be text, image, audio or video. Many studies focus on review the steganalysis algorithm based on steganography and steganalysis classification. This paper will review the image steganalysis techniques based on image type format classification, from image format point of view, focusing on the main and the most commonly used format JPEG, BMP, GIF and PNG.

General Terms
Cover image, stego image, steganography,

Keywords
Steganalysis, Multimedia documents, Image format, frequency domain steganalysis, spatial domain steganalysis.

1. INTRODUCTION
Steganalysis is the art of how to detect the existence of hidden information or hidden message in a multimedia document, and to discriminate stego object and non-stego-object with little or no knowledge about the steganography algorithm. The goal of steganalysis is to collect any evidence about the presence of embedded message [1].

If the steganography is the art of hiding messages into multimedia documents; the steganalysis is the art of detecting such the hidden messages. A message can be hidden in a document only if the content of a document has high redundancy [2].

Although the embedded message changes the characteristics and nature of the document, it is required that these changes are difficult to be identified by an unsuspecting user. On the other hand, steganalysis develops theories, methods and techniques that can be used to detect hidden messages in multimedia documents. The documents without any hidden messages or hidden information are called cover documents and the documents with hidden messages are named stego documents [3][4]. The multimedia document may be text, image, audio or video, this paper will concentrate on the image. Also stego document or stego object in this paper will be the stego image which is the image with hidden message or hidden information, and the cover document or cover object is the cover image which is the original image without and hidden message.

The primary step of steganography and steganalysis process is to identify the image that the secret message will be hidden in, which will called cover image, then use any steganography algorithm to embed the message in the cover image, sometimes by the help of secret key, so the cover image become stego image as shown in Figure 1. After that steganalysis process determines whether that image contains hidden message or not and then try to recover the message from it. In the cryptanalysis it is clear that the intercepted message is encrypted and it certainly contains the hidden message because the message is scrambled. But this may not be true in the case of steganalysis. The stego image may or may not be with hidden message. The steganalysis process starts with a set of unknown information streams. Then the set is reduced with the help of advanced statistical methods [5].

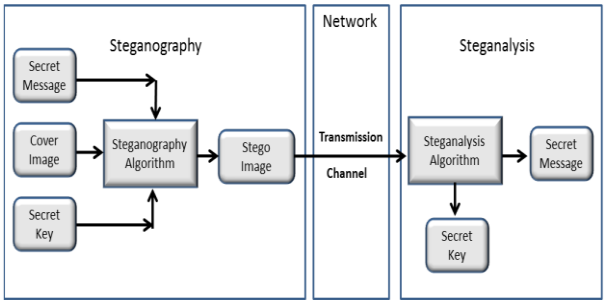

Figure 1: Steganography and Steganalysis Process

Steganalysis can be targeted/specific or blind/generic. In targeted technique, the first look at steganalysis techniques which designed for a particular steganographic embedding algorithm and have a high success rate for detecting the presence of the secret message if the message is hidden with the algorithm for which the techniques are meant for. whereas blind steganalysis is general class of steganalysis techniques which can be implemented with any steganographic embedding algorithm, even an unknown algorithm and produces good results for detecting the presence of a secrete message hidden using new and/or unconventional steganographic algorithms[2][6].

The image steganalysis techniques under both the specific and generic categories are often designed to detect the presence of a secret message and the decoding of the same is considered complementary not mandatory [2].

Steganalysis methods are divided as follows based on the way of detecting the presence of hidden message:

i) Statistical steganalysis

a) Spatial domain

b) Transform domain

ii) Feature based steganalysis

Statistical steganalysis is used to detect the existence of hidden message that is done with the pixels. It is further classified as spatial domain steganalysis (Image Domain) and

Transform domain steganalysis (Frequency Domain steganalysis). In spatial domain, any algorithm is applied directly on the pixels or the pair of pixels is considered and the difference between them is calculated. In transform domain, image transformation and manipulation of algorithm applied, and frequency counts of coefficients are calculated and then histogram analysis is performed. In the other hand, feature based steganalysis used the features of the image to detect hidden message in an image. They can also be used to train classifiers [3].

Also there is a generic steganalysis algorithms, usually referred to as Universal or Blind Steganalysis algorithms, these steganalysis algorithms work well on all known and unknown steganography algorithms. These steganalysis techniques exploit the changes in certain innate features of the cover images when a message is embedded. The focus is on to identify the prominent features of an image that are monotonic and changes statistically as a result of message embedding. The accuracy of the prediction heavily depends on the choice of the right features, which should not vary across images of different varieties. [2].

2. IMAGES FORMATS

Images can be represented in different formats; there are most commonly used formats as: JPEG (Joint Photographic Exchange Group), BMP (Bit Map), GIF (Graphics Interchange Format) and PNG (Portable Network Graphics) [2].

JPEG Image is a popular cover image format used in steganography. The compression method is usually loosely compression, meaning that some visual quality is lost in the compression process and cannot be restored.

BMP are characterized by a lossless LSB plane. BMP files are device-independent files most frequently used in Windows systems, Based on RGB color model. Header region contains info about size and color depth. Data region contains the values of each pixel in a line. LSB embedding on such images causes the flipping of the two gray scale values.

The GIF format Introduced in 1987 by CompuServe, it supports up to 8 bits per pixel and the color of the pixel is referenced from a palette table of up to 256 distinct colors mapped to the 24-bit RGB color space. LSB embedding of a GIF image changes the 24- bit RGB value of a pixel and this could bring about a change in the palette color (among the 256 distinct colors) of the pixel [2]. Easy to decode and display. Images are compressed 20-25% of original size with no loss of quality: data compression only. Uses LZW compression algorithm

PNG Introduced in 1999. PNG supports three main image types: true color, gray scale and palette-based (``8-bit"). JPEG only supports the first two; GIF only the third (although it can fake gray scale by using a grey palette). PNG supports variable alpha channels (transparency).

In images, there are different compression algorithm is available such as “Lossy” Compression is to reduce the amount of information to be transmitted. This is done by compressing the information by permanently losing some of it, particularly redundant information. JPEG (JOINT PHOTOGRAPHIC EXPERTS GROUP) is the image format that follows Loosely Compression [7].

On the other hand, “Lossless” Compression never discards any information from the target image. All the information can be restored even after the image is decompressed.GIF (GRAPHICAL INTERCHANGE FORMAT) and BMP (BIT MAP FILE) are image format that follows Lossless Compression. Importance of Compression is that it helps us to choose the suitable technique to follow [7].

Image steganalysis algorithms can be classified into two broad categories: Specific and Generic. The Specific steganalysis algorithms are based on the format of the digital image (e.g. GIF, BMP and JPEG formats) and depend on the underlying steganography algorithm used. The Generic image steganalysis algorithms work for any underlying steganography algorithm, but require more complex computational and higher-order statistical analysis [2].

When a message is embedded in the image, each of these image formats will have different behavior. Accordingly, there exist different image steganalysis algorithms for each of these image formats. The next section will discuss the recent algorithms for each of these formats.

3. JPEG IMAGES STEGANALYSIS

Joint Photographic Expert Group, also known as JPEG, is the most popular and widely used image format for sharing and storing digital images over the Internet or any PC [8]. It is the common format of the images produced by digital cameras, scanners, and other photographic image capture devices [9]. The popularity of JPEG is due to its high compression ratio with good visual image quality. The file format defined by JPEG stores data in JFIF (JPEG File Interchange Format), which uses lossy compression along with Huffman entropy coding to encode blocks of pixels[8]. Since JPEG images have been widely used in our daily life, the steganalysis for JPEG images becomes very important and significant [10].

Therefore, hiding secret information into JPEG images may provide better camouflage. Most of the steganographic schemes embed data into the non-zero alternate current (AC) discrete cosine transform (DCT) coefficients of JPEG images. As a result, the embedding rate of JPEG steganographic is often evaluated in bit per non-zero AC DCT coefficient (bpac) [9].

There are different file formats available in steganography transform domain but JPEG file format is most popular among the others. The reason is that the size of the JPEG image is very small [7].

Two well-known Steganography algorithms for hiding secret messages in JPEG images are: the F5 algorithm and Outguess algorithm. The F5 algorithm uses matrix embedding to embed bits in the DCT (Discrete Cosine Transform) coefficients in order to minimize the number of changes to a message [2].

3.1 Steganalysis Techniques for JPEG Images

A new universal steganalysis method based on sparse representation is proposed in [10] to overcome the shortage of traditional classifiers in universal steganalysis for JPEG images. Sparse representation, just as its name implies, is to convey the main body of information with as little information as possible, thus simplifying the solving process of information processing. Comparing with the universal steganalysis method for JPEG stego images based on SVM or other classifiers for the training and detection processes, the experimental results prove that the new method has high detection accuracy, and can avoid the “over-fitting” problem of traditional classifiers. Experimental results also prove that the new method is more robust than SVM when the detection images meet with Gaussian noises or Salt Pepper noise.

In [11] Qingzhong Liu et al. proposed a new scheme is for steganalysis of JPEG images, as they see that JPEG images being the most common image format, is it widely used for steganography purposes as there are many free or commercial tools for producing steganography using JPEG covers.

First, they proposed the inter-block of the discrete cosine transform (DCT) and the approximation subband of discrete wavelet transform (DWT). Also they extracted the features on the joint distributions of the transform coefficients and the features on the polynomial fitting errors of the histogram of the DCT coefficients. All features are called original ExPanded Features (EPF). Next, they extracted the EPF features from the calibrated version; these are called reference EPF features. They calculated the difference between the original and the reference EPF features and then merged the original EPF features and the difference to form the feature vector for classification.

They used the feature selection method of support vector machine recursive feature elimination (SVM-RFE) to handle the large number of developed features and a method of multi-class support vector machine recursive feature elimination (MSVM-RFE) to select features for binary classification and multi-class classification, respectively. Then they applied support vector machines to the selected features for detecting stego-images.

This new approach improves the detection performance on several JPEG-based steganographic systems, including JPHS, CryptoBola, F5, Steghide, and Model based steganography, for both multi-class and binary classifications as the experimental results show.

In [12] N.V.S. Sree Rathna Lakshmi proposed a novel steganalytic algorithm which can differentiate between the normal and the stego image, in order to break a steganographic method, whose stego image is of high quality. Her proposed steganographic method is employed to hide multiple images, in a single JPEG cover image without any quality loss as shown in experimental results. Also, both the JPEG and BMP images can be embedded in a JPEG image. A stego image is obtained, after embedding all the images. The steganographic method makes sense for JPEG, PNG and BMP image formats. The steganalytic algorithm exploits '3-Level DWT' and the energy value is calculated, based on which classification between the stego and the normal image is carried out. She proves the accuracy of her proposed algorithm by identifying between the stego and normal image with 90% accuracy, over the existing method which shows 80% accuracy rate as shown in experimental results. The accuracy rate remains stable when different sets of images are tested. The PSNR value of the steganographic algorithm which obtains the stego images are ranges in 51. SVM classifier is employed here to classify between the images. Also, this project can be extended to other transforms such as contour let, Curve let and wavelet lifting.

Seongho Cho et al. in [13] differentiate a stego image from its cover image based on steganalysis of decomposed image blocks. They classify image blocks into multiple classes after image decomposition into smaller blocks, and find a classifier for each class. Then, steganalysis of the whole image can be obtained by integrating results of all image blocks via decision fusion. They conduct Extensive performance evaluation of block-based image steganalysis. There exists a trade-off between the block size and the block number for a given test image. Also they propose to use overlapping blocks to improve the steganalysis performance. Additional performance improvement can be achieved using different decision fusion schemes and different classifiers. They point out that the choice of a proper classifier plays an important role in improving detection accuracy, and show that both the logistic classifier and the Fisher linear discriminant classifier outperform the linear Bayes classifier by a significant margin. The experimental results shown that the proposed method offers a significant improvement in detection accuracy when compared to prior art using a frame based approach. Also block-based image steganalysis offers decision reliability information even with only one test image given, which is not available with the frame-based approach.

Mansour Sheikhan et al. in [14] proposed a blind color image steganalyzer, in which the features are extracted from Contour let domain. The used features come from the statistical features of Contour let coefficients and co-occurrence metrics of sub band images. To evaluate their proposed steganalysis method, they use some popular steganography methods such as OutGuess, JPHS, Model-based and Jsteg with payloads of 10% to 25%. Also they used the Analysis of Variance (ANOVA) method to reduce the number of features; they extracted 576 features per color. The number of features was reduced by ANOVA feature selection technique to reduce the computational load and also improve the accuracy rate of classification, and then the selected features are fed to nonlinear Support Vector Machine (SVM) for classification into stego and clean images.

Experimental results show high sensitivity of contour let and co-occurrence matrix features to data hiding. The Experimental results showed that High sensitivity of features to data hiding and blind steganalysis method achieved a high accurate rate for low embedding rates.

Manu Devi and Nidhi Sharma in [4] presented four different types of steganalysis techniques. These steganalysis techniques are developed against the steganographic methods that use binary images (black and white) as a cover image for a secret message. Unlike gray scale and color images, binary images have a rather modest statistical nature. This makes it difficult to apply the existing steganalysis directly on the binary images.

They work investigate steganalysis that extract information related to a secret message hidden in stego image. Their proposed technique has improved the selected steganalysis techniques by minimizing image-to image variations. They estimate the cover image from the stego image and then compute the difference between the two to minimize the image-to image variation then they extract the feature set from this difference.

Their proposed technique proves its effectiveness as experimental results show. Their proposed method can detect the steganography developed in and estimate the length of the embedded message. They also observed that it is insufficient only using a set of rules to ensure suitable data carrying pixels because the notches and protrusions produced from embedding still can be utilized to mount an attack. They suggest incorporating an adaptive pixel selection mechanism for the identification of suitable data-carrying pixels to alleviate this shortcoming in the steganography. They combined several types of features existing in blind steganalysis techniques for analyzing JPEG images, and applied a feature selection technique for the analysis, which not only improves the detection accuracy, but also reduces the computational resources. An enhancement can be obtained by minimizing the influence of image content as they showed. In

other words, they increased the feature sensitivity with respect to the differences caused by steganographic artifacts, rather than the image content.

In [15], Han Zong et al. proposed a blind JPEG steganalysis method based on inter- and intra-wavelet subband correlations in the wavelet domain to detect the presence of information in a stego image more reliably. Firstly, they calculate the joint probability density of each subband's difference from neighboring coefficients in the horizontal, vertical, and diagonal directions after two-level wavelet decomposition, and then they extract the entropy and energy as features from the joint probability density matrix. Then they decompose the image is into three subbands, and the PDF (probability density function) is extracted from each subband's wavelet coefficient. Finally, they combine the three kinds of features described above to detect the image.

They compared their proposed method with various other blind steganalysis methods, and discussed the impacts of different feature combinations on detection accuracy. As shown in the experimental results, the typical JPEG image stego algorithms such as F5, Jsteg, Outguess, and Jphide show that the proposed algorithm has much better detection accuracy than typical existing blind steganalysis algorithms and offers some detection capabilities for double-compressed images.

Their proposed blind JPEG steganalysis method is based on the correlation of inter- and intra-wavelet subbands in the wavelet domain and feature extraction from the co-occurrence matrix. First, after two-order wavelet decomposition, the joint probability density of each subband's difference coefficients with adjoining coefficients in the horizontal, vertical, and diagonal directions is calculated, and the entropy and energy are extracted from the joint probability density matrix as features. Then, the image is decomposed into three subbands, and the PDF is extracted from each subband's wavelet coefficient. Finally, the three features are combined to detect the image. The features chosen here, entropy, energy, and the combination of PDF moments, had better detection performance than the other sets of features studied.

Souvik Bhattacharyya, Gautam Sanyal in [16] proposed a new approach of LSB steganalysis. They used an auto-regressive model with the help of a SVM classifier to detect the presence of the hidden messages, as well as multiple regression parameters, to predict the relative length of the embedded messages. They carried out Multiple Regression analysis of the cover carrier along with the stego carrier in order to find out the existence of the negligible amount of the secret message. The effectiveness and accuracy of the proposed technique demonstrate in the experimental results.

Their proposed method is also can find the approximate hidden area of the secret message. Also it is a good steganalysis algorithm that is useful for both cases gray-scale and color images which is one of the superiority factors of the proposed method in comparison with the some other existing methods that are efficient either for gray-scale or color image.

In [17] Tao Zhang et al. present a least significant bit (LSB) matching steganography detection method based on statistical modeling of pixel difference distributions. As they notes, the previous research indicates that natural images are highly correlated in a local neighborhood and that the value zero appears most frequently in intensity differences between adjacent pixels. The statistical model of the distribution of pixel difference can be established using the Laplace distribution. LSB matching steganography randomly increases or decreases the pixel value by 1 when the message is embedded; thus, the frequency of occurrence of the value zero in pixel differences changes most dramatically during message embedding. Based on the Laplacian model of pixel difference distributions, they proposes a method to estimate the number of the zero difference value using the number of non-zero difference values from stego-images and uses the relative estimation error between the estimated and actual values of the number of the zero difference value as the classification feature. The proposed algorithm is effective in detecting LSB matching steganography and can achieve better detection performance than the local extreme method under most circumstances as the Experimental results show.

Their works begin with a statistical model of the distribution of pixel differences, estimates the number of zero difference values based on the number of non-zero difference values according to the characteristics of LSB matching steganography, and uses the estimated error as the distinguishing feature for steganography classification, through which we achieve good classification. In our future research, we intend to start with the statistical model of image pixel difference and study the algorithm for estimating the embedding rate of LSB matching steganography.

The paper in [18] presented a learning-based steganalytic method to detect LSB matching steganography in gray images. The paper considers the detection of spatial domain least significant bit (LSB) matching steganography in gray images. Natural images hold some inherent properties, such as histogram, dependence between neighboring pixels, and dependence among pixels that are not adjacent to each other. These properties are likely to be disturbed by LSB matching. Histogram will become smoother after LSB matching, and the two kinds of dependence will be weakened by the message embedding. Accordingly, at the first three features are extracted, which are respectively based on, image histogram, neighborhood degree histogram and run-length histogram. Then, support vector machine is used to learn and discriminate the difference of features between cover and stego images. Three features are extracted to form a joint feature set to train SVM classifiers for detection purpose. The proposed method possesses reliable detection ability and outperforms the two previous state-of-the-art methods as proved in experimental results. Ending by analyzing the individual performance of the three features and their fused feature, the individual performances of the three features are compared. The presented method is by no means optimal.

3.2 Results for JPEG Images steganalysis

The results of the previous section will be summarized in the following tables. All these results are taken from the authors' own experiment results. The values in the table represent the detection accuracy rate of the authors' algorithms or techniques. The value is taken as a maximum value if the author uses different steganalysis algorithms or different classifiers in their experiments, and the value is taken as an average if the author uses different payload. The detection accuracy rate for the colored images will be in table and the gray-scale one will be in another table. Table 1 will show the detection accuracy rate for JPEG colored images.

Table 1: Detection accuracy rate for colored JPEG images

Reference	Ref[10]	Ref[11]	Ref[12]	Ref[13]	Ref[14]	Ref[15]
Detection accuracy	90%	99.90%	90%	95%	96%	97%

This table can be represented in graph to show the ranking of detection accuracy rate for the different steganalysis techniques for colored JPEG images as shown in Figure 2, taking into consideration that each reference used different database, different images numbers, different steganography techniques and the steganalysis technique either its blind or targeted steganalysis techniques.

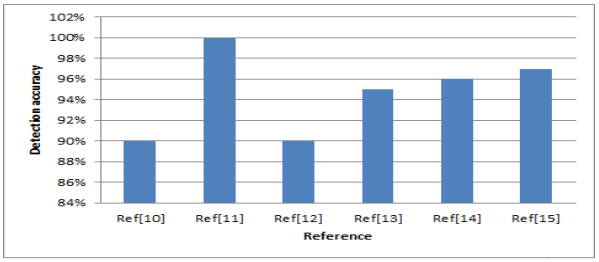

Figure 2: Detection accuracy graph for colored JPEG Images

The detection accuracy rate for gray-scale JPEG images will be shown in Table 2.

Table 2: Detection accuracy rate for gray-scale JPEG images

Reference	Ref[16]	Ref[17]	Ref[18]
Detection accuracy	100%	99.80%	94%

The following figure (Figure 3) represents the detection accuracy graph for gray-scale JPEG images.

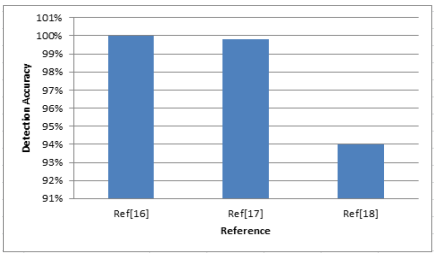

Figure 3: Detection accuracy graph for gray-scale JPEG Images

4. BMP IMAGES STEGANALYSIS

For BMP images The Raw image steganalysis technique is primarily used. The embedding of the hidden message is more likely to result in averaging the frequency of occurrence of the pixels with the two gray-scale values. For example, if a raw image has 20 pixels with one gray-scale value and 40 pixels with the other gray-scale value, then after LSB embedding, the count of the pixels with each of the two gray-scale values is expected to be around 30. This approach based on the assumption that the message length should be comparable to the pixel count in the cover image (for longer messages) or the location of the hidden message should be known (for smaller messages) [2].

4.1 Steganalysis Techniques for BMP Images

Anjani Kumar Verma in [19] uses the supervised learning model for pattern classification such as training data and testing data that classify the stego as well as cover images. In this work, Pre-processed Vectors Diagonal Back Propagation Algorithm (PVDBPA) is implemented to perform the operations which can detect the presence of hidden message and make it with higher detection rate algorithm as a new technique through neural network. Also, BMP Steganalysis using Gray Level Co-Occurrence Matrix has been introduced which also classify data as stego or cover by using feature vectors and analyze through Euclidean distance taken as a measure.

Steganalysis implemented here by using preprocessed vector with BPA. The outputs of the algorithms for one steganographed image have been presented. Secret information is getting retrieved by the proposed algorithms with various degrees of accuracies. It can be noticed that the combined method PVDBPA is giving a newer direction to detecting the presence of hidden information. BPA was trained on 1024 images of group 1 (no hidden message), and 512 images of group 2 (hidden message). Then, patterns from 256 untrained images were computed and given as input to BPA for testing. The work produced a positive classification of 95% and 5% of misclassification.

In [20] H.B.Kekre et al. propose a steganalysis technique for both grayscale and color images based on features that are extracted from co-occurrence matrix of an image. They extract the feature vectors derived from gray level co-occurrence matrix (GLCM) in spatial domain, which is sensitive to data embedding process. This GLCM matrix is derived from an image. They extract 31 features from several combinations of diagonal elements of GLCM. They found a difference between the features of stego and non-stego images and this characteristic is used for steganalysis. They used distance measures for classification like Absolute distance and Euclidean distance. The proposed scheme outperforms the existing steganalysis techniques in attacking LSB steganographic schemes applied to spatial domain as the experimental results show.

They discuss two different distance measures: Absolute and Euclidean which used for the purpose of classification. Euclidean distance gives the best results. It is observed that results obtained using Euclidean distances are better than Absolute distance by around 329% in grayscale images and by 265% in color images. Color images Detection accuracy is better than that of grayscale images by around 18% in Absolute distance and almost same in Euclidean distance. Superiority is observed for low embedding rates. The feature vectors which consist of the diagonal d0 exhibit poor results as compared to feature vectors that do not contain the diagonal d0.

Anuj Kumar and Shivani Khurana in [21] use Adaptive Histogram Equalization Technique that performed on the Stego Image that is category of the Histogram Equalization attack and measuring the quality of Image after applying the Attack. The Adaptive histogram equalization used here as an attack for Steganalysis in spatial domain, that applied on LSB technique to destroy the secret message which is embedded in the cover media. Adaptive histogram equalization (AHE) improves on this by transforming each pixel with a transformation function derived from a neighborhood region.

The difference in image is obtained Here after Steganography using Histogram attack and then they performed the Adaptive Histogram Equalization technique on the Stego Image in which the quality is slightly degraded after Adaptive histogram equalization attack and image is look like as original that is by preserving the quality of image. Hence the enhanced Image is obtained with destruction of Secret message

In [22] Sunaina Verma et al. proposed a new steganalysis technique on the basis of statistical observations on Difference

Image Histograms (DIH) for the reliable detection of classical least significant bit (LSB) steganography which measures the weak correlation between successive bit planes to construct a classifier for discrimination between stego-images and cover images. This proposed technique has two phases. In the first phase, steganography is applied to hide the secret message in the image. In the second phase, steganalysis is used to detect the presence of hidden message, if exist.

In [23] Hossein Malekmohamadi and Shahrokh Ghaemmaghami propose a method for steganalysis of grayscale images using both spatial and Gabor features. Their work has a basis idea, is to use Gabor filter coefficients and statistics of the gray level co-occurrence matrix of images to train a support vector machine. This feature set works well in steganalysis of grayscale images steganographied by LSB matching and Stools as they show. The Gabor filter gives the capability of multi-resolution analysis. Gabor filter is a linear filter and performs like the processes in the primary visual cortex. The multi-resolution structure of Gabor filters is similar to that of wavelets, but without formation of a basis. It has been applied to both face detection and iris recognition. They employed spatial relationships between pixels of clean and altered images to obtain features that are used to train a SVM classifier. Also they used Gabor filter coefficients to construct their input vectors for training an agent. First and higher order statistics from the whole image and its DCT transform have been employed. They achieved high detection rates for very low embedding rates. The extracted features work well in classifying clean images from images steganographied by LSB based algorithms as experimental results show.

The paper in [24] proposes a method for the detection of the Least Significant Bit (LSB) steganography using images as cover objects. Steganography methods attempt to insert data in multimedia signals in an undetectable manner. But these methods distort the underlying signal characteristics, thus allowing detection under careful steganalysis. When embedding is carried out repeatedly, distortion in signal characteristics is highest during the first embedding and decreases subsequently. This principle is used to detect the presence of hidden messages in 24 bit bmp images. This paper proposes improvement to close color pair analysis with stego-sensitive threshold to detect stego objects. Stego-sensitive threshold is calculated based on the structural similarity index measure of the samples. This threshold can be further varied in a range to improve the detection rate. This approach is able to distinguish clean and stego images with promising accuracy as experimental results show.

Experimental results indicate that it is possible to reliably detect the presence of secret message embedded in uncompressed color images using LSB insertion. Based on the hypothesis that on repeatedly embedding message in the image, the changes in the image characteristics is the highest for the first embedding and decreases subsequently. The reliability depends on selection of threshold which is an open ended problem. Variable threshold improves the detection rate. In this paper variable threshold is obtained based on the Structural Similarity Index Measure. After obtaining the threshold, if it is further decreased or increased, then it improves the detection rate of stego images as experimental results prove.

Ziwen Sun et al. in [25] proposed a universal steganalysis method based on co-occurrence matrix of differential image. They calculated the forward difference in three directions, horizontal, vertical and diagonal, towards adjacent pixels to obtain three-directional differential images for a natural image. Then they thresholded the differential images with a pre-set threshold to remove the redundant information. They used the co-occurrence matrixes of thresholded differential images as features for steganalysis. Then they applied Support vector machines (SVM) with RBF kernel as classifier to distinguish between stego images and cover images. The proposed steganalysis method is effective in attacking the popular steganographic schemes applied in spatial domain as proved in the experimental results. The proposed method achieves a detection rate above 95% as the data embedding rate is 0.3 bpp (bit per pixel).

Saeid Fazli and Maryam Zolfaghari-Nejad In [26] propose a new method for steganalysis of grey-scale images as the feature-based classification to devise a blind detector specific to grey-scale images; they developed a new targeted steganalysis method based on GLCM features for grey-scale images. They analyzed the effect of various steganographic processes on the statistical properties of the image. Then they extracted the optimal features from the images, which have high ability in make difference between two groups of cover and stego images. Each feature is calculated in DWT coefficients. They used high order statistics in discrete wavelet transform (DWT) coefficients. After that, they ably a pre-processing of principal component analysis (PCA) on extracted features. They used support vector machines (SVM) to classify image segments into stego or cover cases. The proposed method with comprehensive look into current steganographic techniques in DWT domain is able to detect the presence of hidden messages with more than 90% accuracy in different embedded rates. One of The most important advantages is calculating the GLCM features in the DWT domain that it caused more accuracy and lower error. They applied the detection to several current steganographic schemes.

In [27] Sedighe Ghanbari et al. proposed a new algorithm for Steganalysis using GLCM matrix properties; they investigate some different values in the GLCM of the cover and stego images, they extract features that are different between these images. In their proposed algorithm, they use a combined method of steganography based on both location and conversion to hide the information in the original image and call it image-steg1 image. Then, they hide the information in imagesteg1 again and call it image-steg2. They investigate some different features in the GLCM of the original image and stego images by using GLCM matrix properties. They extract features that are different between these images. These features are used for training neural network and the classification step was accomplished using four layers Multi-Layer Perceptron (MLP) neural network. The algorithm tested on 800 standard image databases and 80% of stego images are detected. Therefore, the proposed algorithm efficiency is 80%.

In paper [28] an improved method for attacking the LSB (Least Significant Bit) matching steganography proposed. In LSB matching steganography the least two or more significant bit-planes of the cover image would be changed during the embedding and thus the pairs of values do not exist in stego image. The proposed method obtained an image by combining the least three significant bit-planes and is divided into 3x3 overlapped sub-images. The sub-images are grouped into eight types, i.e. T1, T2, T3, T4, T5, T6, T7 and T8 according to the count of gray levels. A message is embedded by LSB matching. The alteration rate of the number of elements belonging to T1 is computed. It is found that normally the alteration rate is higher in cover image than in the

corresponding stego image. This is used as the discrimination rule in the method. As the experimental results show, the proposed method gives better results than the previous work which takes into consideration image formed by combination of least two significant bit-planes. Percentage change in detection from two planes to three planes is around 20% in case of cover images. In case of stego images the percentage change in detection from two planes to three planes is up to 15 %. It is observed that percentage detection is more in case of cover images as compared to that of stego images which gives additional information about information hiding. The performance is tested on a database of uncompressed gray scale images. It is an efficient method to detect the LSB matching steganography in gray scale images.

4.2 Results for BMP Images steganalysis

The previous results will be summarized in the following tables. Table 3 shows the detection accuracy rate for BMP colored images.

Table 3: Detection accuracy rate for colored BMP images

Reference	Ref[20]	Ref[24]	Ref[27]
Detection accuracy	97%	62.00%	80%

This table can be represented in graph to see the ranking of detection accuracy rate for the different steganalysis techniques for colored BMP images as shown in Figure 4, also taking into consideration that each reference used different database, different images numbers and different steganography techniques and the steganalysis technique either its blind or targeted steganalysis techniques.

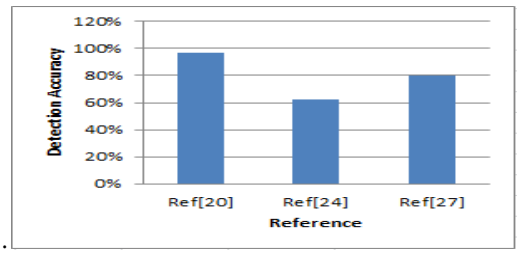

Figure 4: Detection accuracy graph for colored BMP Images

The detection accuracy rate for gray-scale BMP images will be shown in Table 4.

Table 4: Detection accuracy rate for gray-scale BMP images

Reference	Ref[19]	Ref[20]	Ref[23]	Ref[24]	Ref[25]	Ref[26]	Ref[28]
Detection accuracy	95%	99%	92%	62%	84%	98%	65%

The previous table will be represented in figure 5.

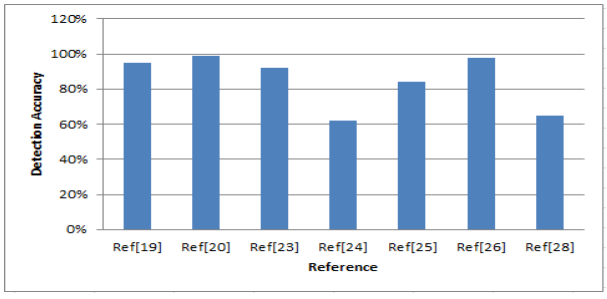

Figure 5: Detection accuracy graph for gray-scale BMP Images

5. GIF IMAGES STEGANALYSIS

The graphics inter-change format (GIF) image is one of the most suitable cover due to its popularity in multimedia and Internet application. Many steganographic methods for GIF images were proposed. However, little attention has been paid to the steganalysis for GIF images [29]. Palette image steganalysis is primarily used for GIF images. The strength of the steganographic algorithm lies in reducing the probability of a change in the palette color of the pixel and in minimizing the visible distortion that embedding of the secret image can potentially introduce. The steganalysis of a GIF stego image is conducted by performing a statistical analysis of the palette table vis-à-vis the image and the detection is made when there is an appreciable increase in entropy (a measure of the variation in the palette colors). The change in entropy is maximal when the embedded message is of maximum length [2].

5.1 Steganalysis Techniques for GIF Images

Hong Zhao et al. in [30] propose a novel blind steganalysis algorithm for palette-Based images, this algorithm is for detecting GIF steganographic. First, they constructed generalized difference images between adjacent pixels, and then extracted the moments of characteristic functions of difference images histograms as features. Second, they used color correlogram technique to capture the global distribution of local spatial correlation of colors in order to measure the dependencies of neighboring colors. They extracted the center of mass of the characteristic function of color correlogram and the absolute moments of auto correlogram, and then classified total of 13 dimension features with machine learning technique. After set of experiments on several existing GIF steganography algorithms the proposed scheme prove its efficiency and gets good performance, especially when the embedding rate is not less than 20%. The average accuracy of their proposed scheme for different GIF steganography algorithms outperforms Lyu's algorithm more than 20% as the experimental results show. It also showed that the proposed scheme achieved similar performance with Fridrich's scheme and higher accuracies comparing to Du's algorithm and biologically inspired features.

They extract the steganalytic features with high efficiency for steganalysis of GIF images while maintaining a low dimensionality. The features generated from gen-realized difference images and color correlogram are highly effective in discriminating stego images from cover images, since they are designed purposely to capture the correlations between adjacent pixels. Combining the steganalytic features with a two-class classification scheme, we differentiate cover images and stego images embedded by several current GIF steganographic algorithms with high accuracy when the embedding rate is no less than 20%. Experimental results also show that the performance of the proposed scheme outperforms most relative works. It also achieves similar performance with Fridrich's scheme. The proposed features outperform biologically inspired features as the experimental results show.

In [29] Rui Gong, HongxiaWang proposed A steganalysis algorithm based on colors-gradient co-occurrence matrix (CGCM). CGCM is constructed with colors matrix and gradient matrix of the GIF image, and 27 dimensional statistical features of CGCM, which are sensitive to the color-correlation between adjacent pixels and the breaking of image texture, are extracted. They give 27-dimensional statistical

features to support vector machine (SVM) technique to detect hidden message in GIF images.

The proposed algorithm is more effective than Zhao's algorithm in [30] for several existing GIF steganographic algorithms and steganography tools, especially for multi-bit assignment (MBA) steganography and EzStego as experimental results indicate. In addition to that, the time efficiency of the proposed algorithm is much higher than Zhao's algorithm. The concept of CGCM and a steganalysis algorithm based on CGCM which proposed is more effective and requires less computing time than the algorithm in [30] for five typical steganographic methods.

5.2 Results for GIF Images steganalysis

The previous results will be summarized in the following table. Table 5 shows the detection accuracy rate for GIF images.

Table 5: Detection accuracy rate for GIF images

Reference	Ref[29]	Ref[30]
Detection accuracy	85%	77%

This table can be represented in Figure 6 see the ranking of detection accuracy rate for the different steganalysis techniques for GIF images.

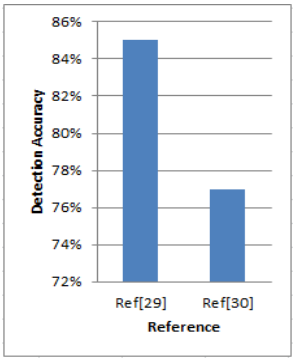

Table 6: Detection accuracy graph for GIF images

6. PNG IMAGES STEGANALYSIS TECHNIQUES

Not enough researches worked in PNG images steganalysis. P.Thiyagarajan et al. in [31] focus on universal image steganalysis method which uses RGB to HSI color model conversion. They proposed a technique using Color Model conversion for discriminating the cover image and stego image (generated from LSB technique). They tested their proposed Steganalysis method on Stego-image produced by their Stego-Image Generator (SIG) tool. They implemented their proposed Steganalysis method along with Stego-Image Generator (SIG) tool using MATLAB 7 and taken bmp and png images as input. The developed Universal Steganalysis algorithm is tested in stego-image database which is obtained by implementing various RGB Least Significant Bit Steganographic algorithms. The Proposed Steganalysis using color Model has been tested with their Image Database and the results were affirmative.

When concluding the Major Steps involved in Steganalysis using HSI Color Model is to convert the given image into HSI Color Model Image, perceive the image for any abrupt changes in the color, and if there is any color distortion in the HSI color image then it is stego image and if there is no color distortion in the HSI color image then the given image is cover image.

Though their proposed method is generic, they tested their method only for stego-images generated by LSB Steganography algorithm. They need to repeat the experiment with large image database available online and to test with other different Steganography algorithms.

7. CONCLUSION

This paper introduce a review on steganalysis techniques from image format point of view, taking into consideration the most commonly used image format, JPEG, BMP, GIF and PNG. Most techniques for JPEG images are frequency domain steganalysis, due to its loosely compression, its steganography techniques is in frequency domain to avoid the loss of the hidden data. In the other hand, most steganalysis techniques for BMP, GIF and PNG images are spatial domain steganalysis. Frequency domain is more robust when compared to the spatial domain.

JPEG image steganalysis is more complex than the other image format because working in frequency domain is not trivial, so it needs more statistical analysis.

Steganography in grayscale images is undetectable by the human visual system, but its steganalysis techniques are easier than steganalysis techniques for colored images. By referring to the results of the pervious literature experiments, grayscale images gives higher detection accuracy rate than the colored images.

Most steganalysis techniques is done for JPEG and BMP images, but BMP images steganalysis needs more effort in slandered way. Most of the BMP steganalysis techniques did not use slandered database in their experiments, they create their own database. Little effort has been done to the steganalysis for GIF and PNG images, much more attention must be paid for GIF and PNG images steganalysis.

8. REFERENCES

[1] Manveer Kaur1, Gagandeep Kaur2, 2014," Review of Various Steganalysis Techniques", International Journal of Computer Science and Information Technologies, Vol.5 (2).

[2] Natarajan Meghanathan, Lopamudra Nayak, January 2010, "STEGANALYSIS ALGORITHMS FOR DETECTING THE HIDDEN INFORMATION IN IMAGE, AUDIO AND VIDEO COVER MEDIA", International Journal of Network Security & Its Application (IJNSA).

[3] Dr.P.Sujatha, Dr.S.Purushothaman, P.Rajeswari, January 2014, "Computational Complexity Evaluation of ANN Algorithms for Image Steganalysis", International Journal of Latest Trends in Engineering and Technology (IJLTET), Vol. 3 Issue 3.

[4] Manu Devi, Ms. Nidhi Sharma, August, 2013, "Improvements of Steganography Parameter in Binary Images and JPEG Images against Steganalysis", INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY.

[5] Krati vyas1, B.L.Pal, January 2014 , "A PROPOSED METHOD IN IMAGE STEGANOGRAPHY TO IMPROVE IMAGE QUALITY WITH LSB TECHNIQUE", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 3, Issue 1.

[6] Dr. Monisha Sharma and Mrs. Swagota Bera, June 2012, "A REVIEW ON BLIND STILL IMAGE STEGANALYSIS TECHNIQUES USING FEATURES EXTRACTION AND PATTERN CLASSIFICATION METHOD." International Journal of Computer Science, Engineering and Information Technology (IJCSEIT).

[7] R.Poornima1 and R.J.Iswarya, February 2013, "AN OVERVIEW OF DIGITAL IMAGE STEGANOGRAPHY", International Journal of Computer Science & Engineering Survey (IJCSES). Vol.4, No.1.

[8] MAHENDRA KUMAR, 2011, "STEGANOGRAPHY AND STEGANALYSIS OF JOINT PICTURE EXPERT GROUP (JPEG) IMAGES", DOCTOR OF PHILOSOPHY DISSERTATION, UNIVERSITY OF FLORIDA.

[9] Bin Li, Junhui He, Jiwu Huang, Yun Qing Shi, April 2011, "A Survey on Image Steganography and Steganalysis", Journal of Information Hiding and Multimedia Signal Processing, Ubiquitous International Volume 2, Number 2.

[10] Zhuang Zhang, Donghui Hu, Yang Yang, Bin Su, 2013, "A Universal Digital Image Steganalysis Method based on Sparse Representation", Ninth International Conference on Computational Intelligence and Security.

[11] Qingzhong Liu, Andrew H. Sung, Mengyu Qiao, Zhongxue Chen, Bernardete Ribeiro, 2010 "An improved approach to steganalysis of JPEG images", Information Sciences 180, 1643–1655, EL SEVIER,

[12] N.V.S. Sree Rathna Lakshmi, May 15, 2014, "A Novel Steganalytic Algorithm based on III Level DWT with Energy as Feature", Research Journal of Applied Sciences, Engineering and Technology, Maxwell Scientific Organization.

[13] Seongho Cho, Byung-Ho Cha, Martin Gawecki, C.-C. Jay Kuo, 2013, "Block-based image steganalysis: Algorithm and performance evaluation", J. Vis. Commun. Image R., ELSEVIER.

[14] Mansour Sheikhan, M. Shahram Moin, Mansoureh Pezhmanpour, 2010, "Blind Image Steganalysis via Joint Co-occurrence Matrix and Statistical Moments of Contourlet Transform", 10th International Conference on Intelligent Systems Design and Applications, IEEE.

[15] Han Zong, Fen-lin Liu, Xiang-yang Luo, 2012, "Blind image steganalysis based on wavelet coefficient correlation", Digital Investigation.

[16] Souvik Bhattacharyya, Gautam Sanyal, AUGUST 2011, "Steganalysis of LSB Image Steganography using Multiple Regression and Auto Regressive (AR) Model, Int. J. Comp. Tech. Appl (IJCTA), vol 2.

[17] Tao Zhang, Wenxiang Li, Yan Zhang, Ergong Zheng, Xijian Ping, 2010, "Steganalysis of LSB matching based on statistical modeling of pixel difference distributions", Information Sciences, Elsevier Inc.

[18] Zhihua XIA, Lincong YANG, Xingming SUN, Wei LIANG, Decai SUN, Zhiqiang RUAN, APRIL 2011, "A Learning-Based Steganalytic Method against LSB Matching Steganography", RADIOENGINEERI, VOL. 20, NO. 1.

[19] Anjani Kumar Verma, June 2014, "A Non- Blind Steganalysis Through Neural Network Approach", International Journal of Multidisciplinary Consortium, Volume 1, Issue 1.

[20] H.B.Kekre, A.A.Athawale, S.A.Patki, 2011, "Steganalysis of LSB Embedded Images Using Gray Level Co-Occurrence Matrix", International Journal of Image Processing (IJIP), Volume (5), Issue (1).

[21] Anuj Kumar and Shivani Khurana, March 2014 , "Steganalysis Technique on Gray scale Image by Varying Message Length using Adaptive Histogram Equalization Attack", International Journal of Computer Applications, Volume 90.

[22] Sunaina Verma,Sandeep Sood, Sukhjeet Kaur Ranade, February 2014," Relevance of Steganalysis using DIH on LSB Steganography", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 2.

[23] Hossein Malekmohamadi, Shahrokh Ghaemmaghami, 2009, "STEGANALYSIS OF LSB BASED IMAGE STEGANOGRAPHY USING SPATIAL AND FREQUENCY DOMAIN FEATURES", IEEE.

[24] H.B.Kekre, A.A.Athawale, S.A.Patki, Feb 2011, "IMPROVED STEGANALYSIS OF LSB EMBEDDED COLOR IMAGES BASED ON STEGO-SENSITIVE THRESHOLD CLOSE COLOR PAIR SIGNATURE", International Journal of Engineering Science and Technology (IJEST), Vol. 3 No. 2.

[25] Ziwen Sun, Maomao Hui, Chao Guan, 2008, "Steganalysis Based on Co-occurrence Matrix of Differential Image", International Conference on Intelligent Information Hiding and Multimedia Signal Processing, IEEE computer society, IEEE.

[26] Saeid Fazli, Maryam Zolfaghari-Nejad, March 2012, "A New Steganalysis Method for Steganographic Images on DWT Domain", International Journal of Science and Engineering Investigations, vol. 1, issue 2.

[27] Sedighe Ghanbari, Manije Keshtegary, Najme ghanbari, March 2012, "New Steganalysis Method using Glcm and Neural Network", International Journal of Computer Applications, Volume 42– No.7.

[28] H B Kekre, A A Athawale, S A Patki, 2011, "Improved Steganalysis of LSB Matching Steganography Based on Counting Alteration Rate of the Number of Neighborhood Gray Levels", International Conference and Workshop on Emerging Trends in Technology (ICWET) – TCET, Mumbai, India.

[29] Rui Gong, HongxiaWang, 2012, "Steganalysis for GIF images based on colors-gradient co-occurrence matrix", Optics Communication, ELSVIER.

[30] Hong Zhao, Hongxia Wang, Muhammad Khurram Khan, 2011,"Steganalysis for palette-based images using generalized difference image and color correlogram", ELSEVIER, Signal Processing.

[31] P.Thiyagarajan, G.Aghila, V. Prasanna Venkatesan, December 2011, "STEGANALYSIS USING COLOUR MODEL CONVERSION", Signal & Image Processing: An International Journal (SIPIJ), Vol.2, No.4.