International Journal of Engineering & Technology, 7 (4) (2018) 4008-4023

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET
doi: 10.14419/ijet. v7i4.16004
Research paper

An extensive analysis and conduct comparative based on statistical attach of LSB substitution and LSB matching

Mohammed Mahdi Hashim [1, 3] *, Mohd Shafry Mohd Rahim [1, 2], Fadil Abass Johi [4], Mustafa Sabah Taha [1, 5], Ali A. Al-Wan [6], Nilam Nur Amir Sjarif [6]

[1] School of Computing, Facility of Engineering, University Technology Malaysia, Johor Bahru, Malaysia
[2] UTM (IRDA) Digital Media Center, Faculty of Computing, University Technology Malaysia, Johor Bahru, Malaysia
[3] Faculty of Engineering, Uruk University, Baghdad, Iraq
[4] Missan Oil Company ,Ministry of Oil, Iraq
[5] Basrah Oil Training Institute ,Ministry of Oil, Iraq
[6] Razak Faculty of Technology & Informatics
**Corresponding author E-mail: comp.mmh@gmail.com*

Abstract

Steganography and steganalysis are the two diverse sides of the same coin, as steganalysis is a countermeasure to steganography. The major function of steganalysis is to differentiate between actual media and suspected media that contains concealed messages. Carrying out this task can be difficult for new adaptive steganography, because modifications made as a result of concealed messages is very minimal. De-spite the availability of so many techniques in recent times, some of the oldest and most commonly used technique in the last years is the LSB substitution and matching techniques. The statistical steganalysis in LSB substitution and LSB matching approach for the digital imag-es is being analyzed and discussed extensively in this paper. The major contribution of the paper is the evaluation of methods, by means of analyzing challenges and comparing approved studies, with the intention of unveiling novel directions which have the potentials of provid-ing improved and effective steganalysis approach.

Keywords: *Information Hiding; Image Steganalysis, Steganography; LSB Matching, LSB Substitution.*

1. Introduction

The increased awareness of using communication security on networks, is due to the harmful environment in which electronic connection between two sides exist. Therefore, there is an intense growth in the area of information hiding; this growth is caused by the significance of confidentiality and privacy. Steganography is one of the most significant and widely recognized branches used in communicating data secretly, while steganalysis, is a branch that involves the detection of data which was previously concealed through the use of steganography [1].

Steganalysis aims at gathering adequate proof about a hidden message, and to destroy the security of the carrier of the hidden message, thereby defeating the purpose of steganography. Majority of the steganalysis algorithms, depend on steganography algorithm which introduces statistical variation between the stego works and the cover [2], [4]. The application of steganalysis has been in the areas of cyber warfare, tracking of illegal activities over the internet and collecting proof for investigations, especially in case of anti-social elements [3]. In addition to the antisocial and law enforcement importance of steganalysis, it is smoothly applied in improving the security of steganography tools through the evaluation and identification of their limitations.

There are different forms of attacks and hidden data analysis. Some of the different forms include; detection, extraction and destroying or disabling concealed data. There is a similarity between attacking cryptographic algorithms and steganography algorithm, because similar techniques are used. The various techniques used for attacking using the availability of the actual cover file, knowledge of the real message, kinds of steganalysis technique and steganography tool have been highlighted by Fabien A.P. Petitcolas as follows: [10].

- Stego only attack – here, it is just the stego object that is accessible for analysis.
- Known cover attack – both the stego object and cover can be analysed.
- Known message attack – a comparison of the message with the stego object can be done, since the message is known.
- Chosen stego attack – analysis can be performed on the stego object and the stego tool (algorithm).
- Chosen message attack – stego-media can be generated from some steganography tool or an algorithm obtained from a known message by the steganalyst. This kind of attack aims at determining matching patterns in the stego-media that can indicate the use of certain steganography algorithms and tools.
- Known stego attack - the steganography tool (algorithm) is known and both the original and stego-object are available.

The cover medium can be a video file, audio file, text file, image file or network packet. The effectiveness of steganalysis increases when there are more elements known to a digital examiner. More so, the progression of steganalysis from just detection to detection and differentiation of embedded message, increases the complexity of steganalysis. This implies progressing from passive to active steganalysis [9].

Several image steganalysis that have been modified are documented and presented. There are two main approaches that have been adopted by scientists. In the first approach, statistical features are extracted from stego and original images. A comparison of these extracted statistical features is then performed, so as to differentiate the original image from the stego image. The second approach is a general approach which employs the use of machine learning techniques. Therefore, the extraction of features is performed on both stego and clean images, a classifier is trained and lastly, the presentation of unseen images is made to the model for evaluation. Some examples of simple classifiers used here, include the artificial neural networks and the Support Vector Machines (SVM). In addition to these two aforementioned approaches, deep learning techniques like deep auto encoders and conventional neural networks have been applied as modern methods. The use of these deep learning techniques enables the automatic extraction and selection of features.

In [14] the review which was performed only covered methods of steganalysis used for jpeg images, while in [13] only methods for universal (blind) detection for image steganography are reviewed. In [11], [12], a different specific and statistical taxonomy of steganalysis was proposed by the authors. All the methods covered in these studies, are old methods. In our review, a comprehensive reference of old and new methods of steganalysis is provided. This includes image streganalysis, particularly in LSB substitution and LSB corresponding approaches, as well as current trends. The performance of the techniques has been evaluated and analyzed using the following metrics; error rate, detection rate and ROC curves in certain embedding rates.

The rest of paper is organized as follows. In Section 2 Overview of steganography. Section 3 analyzed of LSB substitution, while Section 4 analyzed of LSB matching. Comparison of LSB substitution and LSB matching in Section 5. Types of steganalysis approaches is discussed in Section 6, while in Section 7 types of Image based on steganalysis techniques are presented. Section 8 is present the number of related work on attacking LSB algorithms. In section 9 is describes the evaluation on steganalysis. Finally, in section 10, the conclusion derived of this analysis.

2. Overview of steganography

Steganography was first used by the Greeks, when Herodotus wrote messages to the Greeks. It was also used during the period of cold war for security communication in USSR and US. In recent times, new algorithms alongside different media carriers for the protection of confidential information have emerged [95].

Basically, steganography can be described as the process through which hidden messages are embedded in a secretive manner, such that nobody, except that sender and intended receiver(s) can find the messages as illustrated in figure 1. The result will produce a file known as stego object, containing the secrete message in it. There are three major components that make up the basic model of steganography. The cover object is the first component, which is also known as hosting media that conveys the secret message that will be concealed. The second component is a secret message that can be any binary file like image, file or data etc. The third one is a secret key that is utilized in encoding-decoding the concealed message. The output of embedding algorithm is called Stego media or stego object. It is the result obtained after embedding the secret message.

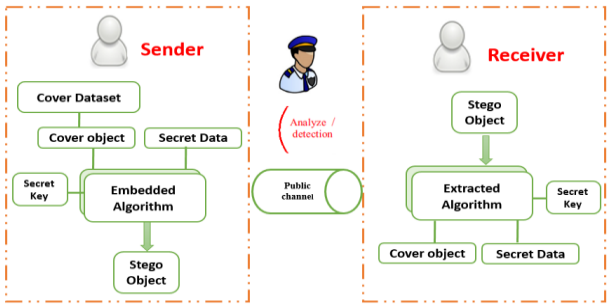

Fig. 1: Illustration Diagram of Steganography Scheme.

The techniques used in embedding steganography can be categorized into transform and spatial domain, which all have different algorithms. The most widely known algorithm is the Least Significant Bit (LSB), and this algorithm can be divided into two schemes known as LSB Substitution and LSB Matching. The following sections contain discussions on the two schemes.

3. LSB substitution steganography

An analysis of the LSB substitution is carried out in this section, according to three perspectives which include the process of embedding, how it affects the intensity histogram and the process of extraction. Later on, in section 6, a detailed description of LSB substitution analysis is given based on various methods.

The LSB of the cover image pixel value is simply replaced by LSB substitution steganography with the value of a single bit of the secret message. The pixel value is left unchanged when there is a correspondence between the LSB values and the bit value of the secret message, while the mismatched LSB is changed by either increasing or decreasing the odd or even pixel values by ne respectively [16], as shown in Figure 2.

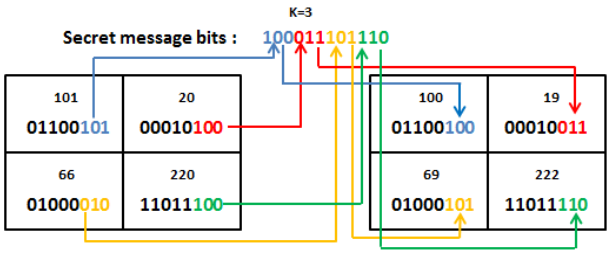

Fig. 2: Possible Pixel Value Transitions with LSB Substitution.

$$P_c = \begin{cases} P_s - 1 & if\ A \neq LSB\ (P_s) \text{ and } P_s \text{ is Even} \\ P_s + 1 & if\ A \neq LSB\ (P_s) \text{ and } P_s \text{ is Odd} \\ P_s & if\ A = LSB\ (P_s) \end{cases} \quad (1)$$

Where, P_c and P_s represent the stego and cover image pixel values respectively, and A is the desired bit value of the secret message. Therefore, the embedding processes of the LSB substitution can be described as follows:

1) for $n = 1, \ldots, i(k)$
2) $lsb = LSB\ (p_n)$
3) if $lsb \neq k_n$ them (2)
4) $lsb \leftarrow k_n$
5) endif
6) end for

Where $i(k)$ contain the message bits. The pixel p_n of the image is first taken, followed by its $LSB\ (p_n)$ value. The lsb of an even number will be 0, while that of an odd number will be 1. We then contrast this with the message bit k_n. If prior to this time they are similar, then no action is required, if otherwise, lsb should be changed with k_n . This process continues even as $i(k)$ is not zero.

In order to perform the analysis of the effect of LSB substitution on cover image intensity histogram, we hypothesize that there is a possibility of 50% for the LSB of the cover image pixel value that already has the desired value. Thus, for an embedding rate of p, the probability of improved pixel values will be (p/2), subsequent to the process of embedding, the unmodified pixel values will be (1-p/2). This implies that changing each embedding message bits, requires 0.5-pixel values. Simply put, it has an embedding efficiency of 2-bits of the secret message per one embedding change. Therefore, an estimation of the intensity histogram could be carried out using equation 2 as follows:

$$h_c(n) = \left(1 - \frac{P}{2}\right) h_s(n) + \frac{P}{4} \begin{cases} h_s(n-1), & n \text{ is Odd} \\ h_s(n+1), & n \text{ is Even} \end{cases} \quad (4)$$

Where n is a grey-scale level which ranges from 0 to 255, h_c and h_s indicate the number of pixels in the stego and cover images respectively, with grey-scale value of n.
Where n denotes a grey-scale level ranging from 0 to 255, h_c and h_s represent the number of pixels in the stego and cover images respectively, with grey-scale value of n.
The result of this kind embedding is an imbalanced distortion and the production of 'Pairs of Values' on the intensity histogram of the stego image. The detection of LSB substitution can be easily carried out by current methods of steganalysis including RS [37], SP [38], and WS, because LSB substitution is naturally asymmetric [45]. The processes of extraction for LSB Substitution Stego-method are as follows:

1) $for\ n = 1, \ldots\ldots, i(s)$
2) $rk_n \leftarrow (s_n)$
3) $end\ for$

The pixels of supposed image are represented by $i(s)$. Run the loop $i(s)$ in place of (k), due to the difference between the processes of embedding and retrieval. The LSB value of each pixel rk is recovered and translated to ASCII, thereby enhancing the clarity of the message and making the message readable to the extent that the embedded message is viewed as claptrap when the LSBs of the image is seen. If the embedded message length is known, then the loop will be ended upon the completion of the message length, and then just the message will be retrieved.

4. LSB matching steganography

In order to perform the analysis of LSB matching steganography, consideration is given to the process of embedding as well as how it affects the intensity of histogram of the cover image. Later, in section 6, a detailed description of LSB matching is provided.
LSB matching which is also referred to as ±1 embedding is a sophisticated version of LSB substitution. This version is proposed to randomly increase or decrease cover image sample value by one for LSB mismatched with secret bit, rather than just substituting the LSB of the cover image [15]. Figure 3 contains the probable pixel value transitions of ±1 embedding.

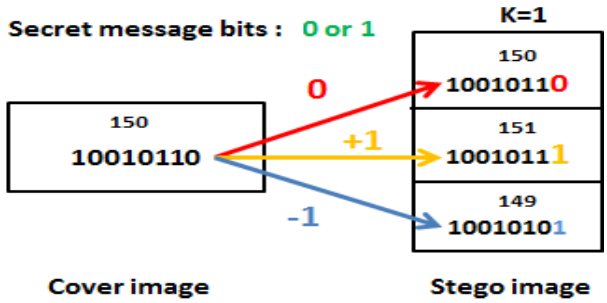

Fig. 3: Possible Pixel Value Transitions with LSB Matching.

The boundary limitation should be maintained by the random increase or decrease in pixel values, and the pixel values should be constantly be between the range of 0 and 255. In other words, no subtraction of 1 should be made from the pixel values of 0 by the process of embedding, neither should 1 be added to pixel values of 255.
The asymmetry changes which can be made to cover image is avoided by the random ±1 change made to the mismatched LSB pixel values; this is the case with LSB replacement. Therefore, the detection of LSB matching is considered as more difficult than LSB replacement [17]. The formal representation of the embedding procedure of LSB matching is given as follows:

$$P_c = \begin{cases} P_s - 1, & if\ A \neq LSB(P_s) \text{ and } (K < 0 \text{ or } P_s = 255) \\ P_s + 1, & if\ A \neq LSB(P_s) \text{ and } (K > 0 \text{ or } P_s = 0) \\ P_s, & if\ A = LSB(P_s) \end{cases} \quad (5)$$

Where the stego and cover image pixel values are denoted by P_c and P_s respectively and is an independent and identically distributed random variable with uniform distribution on (-1, +1).
An embedding rate of P is considered for the intensity histogram. There is a 50% probability that the desired LSB is contained in the clean image pixel, implying that after the process of embedding, change will occur in (P/2) of the cover pixel values. Thus, approximated unmodified pixel values will be (1- P\2), meaning that, 0.5-pixel values are required for the embedding each message bit. Simply put, its embedding efficiency is 2-bits of the secret message per one embedding change. Equation 5 below can be used in obtaining the intensity histogram of the stego image [18]:

$$h_c(n) = \left(1 - \frac{P}{2}\right) h_s(n) + \frac{P}{4} [h_s(n+1) + h_s(n-1)] \quad (6)$$

As earlier mentioned, the asymmetric property can be prevented by the LSB matching from making modifications to the cover image. Nevertheless, [19] claims that reduction to a low pass filtering of the intensity histogram occurs in ±1 embedding. This means that more high-frequency power is contained in the cover histogram than the histogram of the stego image [18], which allows steganalysis to detect the presence of the secret message embedded with LSB matching.

The following describes the embedding processes of the LSB matching:

1- $for\ k = 1, \ldots\ldots, i(n)$
2- $lsb = LSB(p_k)$
3- $if\ lsb \neq n_k\ then$
4- $p_k = p_k + 1\ or\ p_k = p_k - 1\ to\ make\ lsb = n_k$
5- $end\ if$
6- $end\ for$ (7)

The message bits are contained in $i(n)$ C. The pixel p_k of the image is first taken, and its $LSB\ (p_k)$ value. No action is required when corresponding bit and $LSB\ (p_k)$ are now similar, otherwise, there should be increase or decrease in p_k, such that the $LSB\ (p_k)$ becomes the matching bit. This process continues even as is not zero.

5. Accurate comparison of LSB substitution and LSB matching

Many differences and similarities indicated when we compare the LSB substitution and LSB matching techniques according to analysis different studies and methods. The major points of comparison are summarized in table 1.

Table 1: Comparison between the LSB Substitution and LSB Matching

S. No	Comments	LSB Substitution	LSB Matching
1	Easy and simple to implement	✓	✓
2	Spatial domain method.	✓	✓
3	Use the least significant bit (LSB).	✓	✓
4	The same change in statistical properties of the cover image.	✓	✓
5	Embedding rate is 1	✓	✓
6	The rate of alteration is 50%.	✓	✓

7	The secret message can be lost due to the intruder inverts all the LSBs.	✓	✓
8	The secret message can be easily recovered by the unauthorized person.	✓	✓
9	The static method i.e. they sequentially hide the message in LABs of the image pixel value.	✓	✓
10	Same retrieval operation.	✓	✓
11	The process of substitution is performed for replacement of message bit.	✓	
12	Only the LSB of the pixel will change.	✓	
13	In Visual attack, only the LSB bit plane will be changed.	✓	
14	Intrinsically asymmetric, i.e. an even valued pixel will either keep its value or be incremented by one.	✓	
15	The process of adjustment is performed for matching of message bit.		✓
16	More bit modification. There are possibilities that all bits may have change.		✓
17	In Visual attack, least bit plane has the change at great extinct and remaining all bit planes may also have some change.		✓
18	The consequent pixel value is arbitrarily incremented or decremented, thus removing the asymmetry of even and odd pixels.		✓

6. Typical steganalysis approaches

Based on section 2, the secret message in different kinds of media is hidden by steganography in manner that the existence of concealed messages is a secret. Nonetheless, as a result of the amendments in the carrier media, some artifacts which could point the presence of the embedding process, should be provided. The artifacts which are caused by the process of embedding can be observed using various methods, and some of the main methods include, structural, visual and statistical steganalysis.

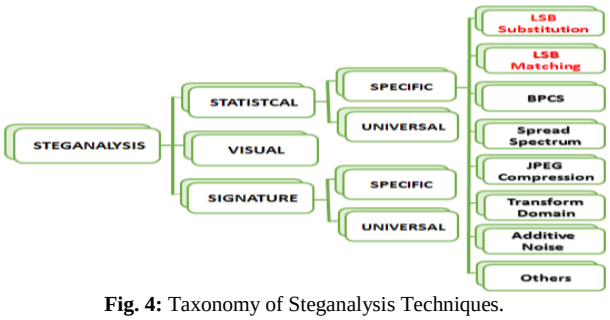

Fig. 4: Taxonomy of Steganalysis Techniques.

6.1. Visual steganalysis

One of the popular ways through which images are attacked, is by exposing the least significant bits of an image and performing an analysis of its randomness using human eyes, since the eyes of humans can perform complex analysis compared to computers [20]. Stego objects, which resemble their cover medium, are created by majority of the steganography algorithms.

It has been proven and clearly stated in [21], that the least important bits of luminance values of digital images are not totally random, even though, several authors have made wrong assumptions that the least significant bits in the image's luminance is completely random. Thus, a completely random noise could enhance the detection of the presence of the hidden message as illustrated in Figure 5 .

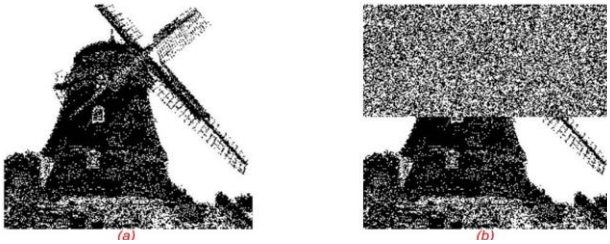

Fig. 5: The Visual Attack; A) Clean Image and B) Stego Image.

Nevertheless, the elimination of unmodified parts of a stego image, makes it possible to notice signs of manipulation. Therefore, the existence of a hidden message may by revealed by a visual attack, if the features of the image that characterize it as stego can be identified using steganalysis.

The commonest kind of visual attack is related to Least Significant Bit (LSB) steganography. After an image has been changed to its binary form, then the retrieval of the LSB plane occurs. The number of even values are as much as odd values in an image, implying that there are approximately as many 1's as there are 0's in its LSB plane. However, the conversion of text to binary, results in more 0' than 1's. This is an indication of a visual inconsistency, and this, enables the forensic examiner to categorize the image as stego. However, the only time this kind of steganalysis technique is effective is if the stego image was produced using inferior steganography algorithm. However, when unaltered parts of a stego image are removed, it is possible to observe signs of manipulation. Hence, if a steganalyst can identify those features of the image that characterize it as stego, a visual attack may reveal the existence of a hidden message.

The most common form of a visual attack concerns Least Significant Bit (LSB) steganography. The image is converted to its binary form and the bits in the LSB plane are retrieved. In an image usually, there are as many even values as there are odd, typically saying that there are approximate as many 1's as there are 0's in its LSB plane. When text is converted to binary, however, there are often more 0's than 1's. This indicates a visual inconsistency and helps the forensic examiner to classify the image as stego. However, this steganalytic technique is successful only when a poor steganography algorithm was used to produce the stego image. Furthermore, hidden messages can be detected using indicators such as increase or decrease in unique colours in stego images, and variation in file size between stego images and cover image.

6.2. Signature steganalysis

This another method of steganography which is used in hiding secret information, while manipulating the images and other digital media, such that the images and digital media are invincible to human eye [22]. Repetitive patterns (signatures) of a steganography software are observed using this kind of steganalysis technique for the purpose of detecting the existence of concealed message. For instance, an addition of the string CDN is always made to the end of the file when an image when a message is embedded; this is done using Hider man steganography software as illustrated in Figure 7. Thus, looking out for repetitive and obvious patterns (signatures) of a steganography tool, is another way of determining the presence of concealed message in suspicious image.

These attacks are particularly related to palette images for LSB embedding in indices to the palette. The attacks are easy and offer promising results when the embedding of message is done sequentially. However, their reliability is low, and it is difficult to automate them. There are different kinds of signature steganalysis methods like [24-26].

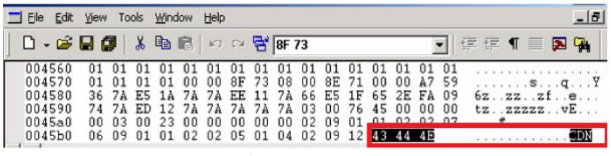

Fig. 6: LSB Plane after Randomized Embedding [23].

6.3. Statistical steganalysis

Statistical steganalysis is described as the techniques which are developed based on the analysis of the procedure of embedding and determining specific statistics that which are altered due to the process of embedding. Thus, in order to achieve maximum accuracy for steganalysis, the process of embedding must be properly understood. A direct application of the steganography algorithm is made on the pixels in spatial domain. In statistical steganalysis, Least Significant Bit Substitution is one of the commonest technique and has been introduced in the current study. There are two approaches that fall under the Least Significant Bit Substitution (LSB), and they are LSB substitution and LSB matching.

7. Types of image based steganalysis

As earlier stated there are two broad classes of steganalysis.

7.1. Specific/targeted steganalysis

The specific steganalysis which is also referred to as targeted steganalysis, is specifically designed to attack a specific kind of steganography algorithm. The statistical trends of the stego image and embedding methods are known to the steganalyst if it is embedded with an algorithm that is known. The effectiveness of this method is higher if tested on images, whose embedding techniques are known. However, if the steganalyst does not have any information about the algorithm, the method may not be effective. There are various methods of steganalysis, and they include [27-30].

7.2. Blind/generic/universal steganalysis

There are several methods that fall into the category of blind steganalysis [31-33]. A less specific class of steganalysis techniques can be designed in a way they can work with any steganography embedding algorithm, even if it is an unknown one. These kinds of techniques are referred to as Universal steganalysis techniques or Blind. More general class of steganalysis techniques independently can be designed to work with any steganography embedding algorithm, even an unknown algorithm. Such techniques have been called as Universal or Blind Steganalysis techniques. These techniques are also regarded as a stronger and modern methods of attacking a stego media, because they have the capability of detecting various kinds of steganography content, even when an algorithm is unknown. However, the specific algorithm which has been used in embedding data cannot be detected using this method, in as much as the training set is not trained using that specific sego algorithm. The method focuses on designing a classifier that relies on the correlations or features that exist within the natural cover images. The most popular and modern methods of universal or blind steganalysis include the extraction of the statistical characteristics also regarded as features from the given set of images. Universal steganalysis is performed based on the following two phases:

1) Feature Extraction.
2) Classification.

7.2.1. Feature Extraction.

This process involves the creation of a set of unique statistical characteristics of an image. These attributes are referred to as features. Feature extraction basically involves a reduction of dimensionality. The features which are extracted must be sensitive to embedding artifacts. Some of the methods of feature extraction include; wavelet decompositions, Markov empirical transition matrix, image quality metrics, moment of image statistic from spatial and frequency domain, co-occurrence matrix and moment of image statistic histograms.

7.2.2. Classification

This phase involves the placing the images into classes based on the values of their features. One of the major classifications of steganalysis is supervised learning, which gives room for learning with some supervision. This kind of learning involves a set of training inputs, which include input features given as input to train the classifier. Subsequent to the training, the given features are used in predicting class label. The following are the most popular kind of classifiers which are used in steganalysis:

- Multivariate regression: It consists of regression coefficient. Minimum mean square error is used here to predicting regression coefficients in the training phase.
- Fisher linear discriminant (FLD): here a linear combination of features is required for the purpose of maximizing separations. This method of classification involves projecting multi-dimensional features into a linear space.
- Support vector machine (SVM): This classification method learns from the given sample. Based on a specific set of features, the machine is trained to identify and allocate class labels.
- Artificial neural network (ANN): this is a model that processes information which is capable of stimulating biological neuron system. It includes the collection of PE such as neuron. When this method of classification is used, feed forward and back propagation neural networks are usually used. There are two steps involved in this method of classification, and they are training and testing. In the first step, which is the training step, the network relates the outputs with specific input patterns, through a modification of the input weights. On the other hand, the testing phase involves the identification of input pattern, and afterwards determining the associated output. In this paper, the presence of hidden information is detected using ANN classifier. hidden information.

8. Various attacks on LSB methods

Westfeld and Pfitzmann were the first to propose the statistical steganalysis. Using this technique helps in the identification of Pairs of Values (POVs) that are exchanged during the process of message embedding. The Pairs of Values could be quantized DCT coefficients, pixel values or palette indices which vary in the LSB. According to these scholars that proposed this technique, the frequencies of each of the two-pixel values in each POV tend to lie far from the mean of the POV. These near-equal POVs in images and subsequently embedded information are detected by the chi-squared attack detects. Messages which have been sequentially embedded can be accurately detected by the Chi-squared method. However, the reliability of detection may reduce when embedding is randomly done [35].

Another method was proposed by Fridrich et al. [36]. The method is used for the detection of LSB embedding in 24-bit colour images. This is known as the Raw Quick Pair (RQP) method. Through the use of, close pairs of colours which were created by LSB embedding. Close color pairs is an indication that two colors vary only at LSB. The number of close color pairs increases when message are being embedded into images. Thus, an image can be characterized as an image or stego image just by counting the number of close clour pairs. Authors showed that even for secret message capacities of 0.1 –0.3 bits per pixel, a high degree of detection accuracy can be achieved. This method is limited by the fact that it cannot be applied to any other image apart from coloured ones.

The limitation of the previous method gave rise to the emergence of a new scheme proposed by Fridrich et al. The new scheme was proposed for the purpose of detecting LSB embedding in both gray-scale and coloured images; this is known as the so-called RS steganalysis [37]. This technique involves dividing image into groups and taking measurement of noise in every group. Subsequently, the LBS of a given set of pixels in each group are flipped using a mask (i.e. the pattern of pixels to flip), and then every group is classified as singular or regular depending on the increase or decrease of the

pixel noise within a group. The classification is repeated for a dual type of flipping.

Based on the work of Fridrich, Dumitrescu et al. presented a generalized case of methods given in [38- 40]. Here, finite state machine, whose states selects multisets of sample pairs referred to as trace multisets. Through the use of this machine, the formulation of a quadratic function through which the length of embedded information can be estimated with high precision is made possible.

Pairs Analysis, which is a method for 8-bit GIF images was proposed by Fridrich et al. [41]. Here, patterns which have been formed by pairs of colours (colour cuts), are used by in estimating the length of a secret message. An entropy-like quantity R is used to take measurements of colour cuts structure. The entropy-like quantity R is a quadratic function of the length of the secret message. The length of the unknown message is estimated from the stego image using this method.

A method known as the Gradient Energy-Flipping (GEFR) was proposed by Li Zhi et al. [42]. This method is used in calculating the gradient energy of both stego and cover image. Afterwards, the message length is estimated using the Gradient Energy curve. The presence of a secret message is accurately detected if the rate of embedding is greater than 0.05 bits per pixel.

Another technique, which is based on the variation in image histogram was proposed by Zhang and Ping [43] for grayscale images. In order to determine a weak correlation between the least significant bit (LSB) plane and other bit planes, the use of translation coefficients between differences in image histograms were used here. A classifier was then constructed using this measure, so as to differentiate the stego-image from the carrier image. These scholars found that the rates of embedding varied from 0% to 100% in 10% increments, while at topmost, the rate of detection reached an average of 96.03%.

In [44] a method of detection, known as least bit (LSB) matching steganography method was presented by Tao Zhang et al. based on statistical modeling of pixel difference distributions. These previous researchers noted that there is a high correlation between natural images within a local neighborhood, and that the frequent appearance of the value zero is often in intensity differences between adjacent pixels. The Laplace distribution can be used in establishing the statistical model of difference in pixel distribution. The value of the pixel is randomly increased or decreased by 1 when the message is embedded in LSB matching steganography. Hence, there is a dramatic change in the frequency of occurrence of the value zero in the differences in pixel distributions. The researchers also proposed a method which can be used in estimating the number of the zero difference value, using the number of non-zero difference values from stego-images. More so, in this method, the relative estimation error between the actual and estimated values of the zero-difference value, is used as the classification feature. Through the use of the proposed algorithm, LSB matching steganography is effectively detected, and better detection performance can be achieved compared to the traditional extreme method in most cases. These have been proven by the results of experiments conducted.

In another study, Ker's WS variant [5] was modified in order to obtain a new version of WS. The new version is proposed for the detection of small payloads exclusively concealed in the least detectable position of a cover. Its performance is then evaluated in comparison to other popular methods. This method is designed to estimate payload sizes which are embedded with naïve adaptivity embedding. Through an experiment, an investigation of the influence of the choice of the adaption criterion is carried out. The adaption criterion is the function which detects spots that are supposedly secured in a heterogenous cover. It was found that, in comparison to our proposed specialized WS method, stronger security is provided by the adaptivity criteria which are difficult to retrieve from the stego image alone [45].

Another new machine-learning framework was proposed by a group of authors for quantitative steganalysis in high-dimensional feature spaces. The authors aimed at extending the feature-based quantitative steganalysis to modern robust models [8, 17] – high-dimensional statistical image descriptors that have been shown to substantially improve classical (binary) steganalysis. When robust models are used for quantitative steganalysis, it is normal to expect high performance. In this framework, the ideas of [23] are combined with that of [19]. In [23], quantitative streganalysis is formulated by the authors as a problem of regression in a suitable feature space, while in [19], an ensemble framework gives room for the use of high-dimensional feature spaces for binary steganalysis. Through the process of gradient boosting, a series of repressors are assembled by the proposed system [11]. The individual base learners are different regression trees, whose splitting criterion is modified to show the particular nature of feature spaces in steganalysis [46].

A combined method for steganalysis of LSB steganography grey-scale images. In order to form the new proposed techniques, two methods in [17], [20] are combined. The technique is developed by merging two output features of two methods, and then using a scheme for dimension reduction such as PCA [47].

A novel methodology to detect information hidden in the LSB plane of a natural raw image.

Within the framework of hypothesis testing theory, the problem of detecting hidden information is cast. Here, the heteroscedastic noise model is expoited, thereby enabling the estimation of noise variance, and improving the performance in terms of detection. One of the criteria considered when data is being hidden using this method is the clipping of picture. It involves the analyses of underexposed and overexposed pixels which are statistically modeled and considered for pixel embedding. The major strength of the proposed approach is the GLRT design that enables the detection of data that have been concealed in clipped images in a way that a prescribed false alarm rate is guaranteed, while high level of detection is ensured. While the impact of clipping phenomenon cannot be tolerated by other detectors, the proposed approach can [48].

A special steganalyzer has been introduced, based on the analysis of the pixel value difference (PVD) histograms of the cover and stego-images. The special steganalyser is proposed based on the least significant bit matching steganalyzer revisited (EA-LSBMR) for the edge adaptive image steganography. The sharper edge regions in the cover images is used by the EA-LSBMR steganography to embed the secret message, thereby achieving a higher level of security. However, there are limitations associated with this method, and this include abnormal increase at some position of the PVD histogram. It is based on the weakness of the EA-LSBMR steganography that the special steganalytic method was designed. Results of detailed experiments showed that the EA-LSBMR steganography can be effectively defeated by the proposed method [49].

A machine-learning based detector, which uses co-occurrences of neighboring noise residuals as features, was proposed by Fridrich et al. The features were adapted by researchers as a means of detecting LSB replacement by making them aware of pixel parity. Afterwards, two new major concepts were introduced by the researchers; parity-aware residuals and calibration by parity. Findings show that when a cover source is known, the building of a binary classifier can be carried out accurately, in comparison to the best WS and structural detectors in both uncompressed images as well as in decompressed JPEGs. This improvement is significant for especially very small rates of change [50].

A new method which is based on the theory of hypothesis testing, is proposed in this paper. The method is specifically designed to enable the detection of hidden data with the LSB matching. This paper proposes a novel method, based on hypothesis testing theory, to detect data hidden with the LSB matching. Using all the parameters of the image, a test through which the detection power is asymptotically maximized, while guaranteeing the probability of false alarm, is presented analytically alongside its statistical properties in a closed form. Through this, an asymptotic upper-bound for the power of any detector is provided for LSB matching. Practically, the parameters of the image are known. A Generalized Likelihood Ratio Test (GLRT) is proposed and its statistical properties are also analytically established [51].

The relationship between the pixels of the image was analyzed by Guo et al. using the matrix of co-occurrence, and some features of

an average co-occurrence matrix were constructed. In this paper, the author proposed a novel LSB matching steganalysis scheme that can be used for gray images. This method exposes the significance between pixels in the LSB matching stego image from the matrix of co-occurrence matrix. High accuracy of close to 100% at high rate of embedding can be achieved using this method. The differences between the cover image and the stego image are strengthened so as to increase the accuracy at low rate of embedding; the aim of this is to improve the performance of the scheme. The extraction of two 8-dimensional feature vectors was carried out independently from the restoration and test images, and subsequently, the steganalysis is carried out by merging 16-dimensional feature vector with FISHER linear classification [52].

Cogranne and Retraint (2013) have recommended the use of the hypothesis testing theory of LSB matching, as they have used in solving problems related to the detection of hidden bits in stego image based on LSB LSB matching scheme [53].

There are three major contributions of this proposed scheme. Firstly, knowing the parameters of the cover medium, results in the establishment of the strongest powerful LRT. An AUMP test which is based on the LRT, and asymptotically increases the detection power, regardless of the rate of embedding the hidden data, is presented. Secondly, an analytic calculation of the detection power of the AUMP is carried out. Through this, the upper bound for the power of detection of any detector can be defined. Thirdly, if the parameters of the inspected medium are not known, then two different local estimations of the unknown parameters are used in proposing [54].

The presence of stego content in images have been detected by Patsakis et al. [55] using compressive sensing. Compressed sensing (CS) which is a growing theory in signal processing, asserts that it is possible to recover any sparse signal from far fewer samples or measurements than that suggested by the traditional Nyquist theory. This theory has been applied in different areas such as image watermarking and steganography.

The authors here first divided the image into equalized and nonlapping blocks. Then each block is modeled as a multivariate Gaussian distribution (MGD), which obviously contains the correlation of the pixels in a block. A novel detector for LSBM was derived through the use of likelihood ratio test and formula derivation. In addition, the paper proposes an improved way of calculating the detector. The new calculation which goes beyond just adding up the detectors among all the blocks within the image, is able to perform Pixel Selection (PS) through an adaptive selection of the blocks with most minimal noise. The detector values within these blocks significantly differ between cover and stego. Results of experiment reveals that the performance of the proposed detector is better than that proposed by previous works [10], as it level of accuracy [56].

In [57] a modified version of the spatial rich model [53] for steganalysis of color images was proposed. Three-dimensional co-occurrences of residuals computed from all three-colour channels was used in the extraction of the extra features which were used. The use of these features can be employed in capturing dependencies across color channels. Three image databases were used in performing experiments, and the include; different versions of BOSSBase v1.01colour and an embedding rate of 0.4 bpp for WOW and 0.1 bpp for LSB Matching. Results obtained from the experiments showed that LSB Matching steganography in images can be detected with high efficiency using the proposed feature set (18,157 features). The average detection error for one payload was found to be 0.0297–0.1790 (LSB Matching for the three test sets), while for different payloads (0.05–0.5 bpc) is also small as contained figures 2 and 3 of the paper.

In this paper, the messages which are embedded using spatial least bit (LSB) matching as independent noises for the cover image, are modelled. It was found that despite the large distance that exist between pixels, the histogram of the variation between pixel gray values is levelled by the stego bits. In this study, by using the characteristics function of difference histogram (DHCF), the authors proved that a decrease occurs in the center of mass of DHCF (DHCF COM) subsequent to the embedding of messages. Therefore, the DHCF COMs are calculated as different features from the pixel pairs with varying distances. An image which is derived through an average operation is used in calibrating the features, which are subsequently used in training a support vector machine (SVM) classifier. Based on the results of the experiments, LSB can be tackled using features which were extracted from the differences between nonadjacent pixels [58].

The steganalysis of LBP based LSB matching was proposed by Xinlu et al. (2014), who embedded big messages into cover image using Least Significant Bit (LSB) matching and steganography methods, obtained results that are statistically and visually unperceivable. However, it was found that the correlation in adjacency of pixels was effected in smooth areas of images as 50% of the payload pixels were modified by 1. The main reasons why the Local binary patterns (LBPs) were initially proposed, were to serve as texture features and to be used in efficiently making summary of local image structures by comparing pixels with their neighbours. The features were trained and classified through the extraction of multi-scaled rotation invariant LBPs as unique features from smooth pixels and linear support vector machine. Results of experiment showed that the method is superior in terms of detecting with higher accuracy [59].

A new method for steganalysis of coloured image was developed by Olguin-Garcia et al., [60], based on histogram characteristic function center of mass detecting histogram changes in each R, G, and B channels. Here, LSB matching steganography method is used in creating stego images. Then, to discover the sufficient threshold, the density function is computed, and afterwards, the values of threshold are determined with various payloads.

In another work, a new universal steganalysis method is proposed for use in both JPEG and spatial domains. The natural statistics of an image can be altered if message is embedded within the media, thereby resulting in weak noises. Thus, the proposed method aimed at extracting the steganalysis features by exploring the disruption of the natural image statistics. The alterations in natural scene statistics were explored using singular value decomposition [61].

The focus of the paper was on improving the efficiency of steganalysis method and at the same time proposing a steganalysis method for the LSB replacement attack. The purpose of the study was to investigate the steganalysis method for LSB replacement, with the aim of refining the detection accuracy as well as minimizing the dimensions of feature vectors, thereby avoiding the problems of dimensionality. At the end of the study, a method for image steganography forensics for LSB replacement was presented. The proposed method involves the decomposition of a grayscale image into eight bit-planes, and calculating the difference in sub-matrix of the bit-planes. The behaviours of the GLCM of the sub-matrix were investigated, and subsequently, significant features were statistically extracted from the GLCM. In order to be able to differentiate the cover images from the stego images, LS-SVM was used as a classifier [62].

In this paper [63], the authors, carried out an investigation to determine if detection function can further be improved. This was done through a synchronization of the training, and testing images by means of configuring the Bayer color filter array and dividing the higher order statistics (co-occurrences) accordingly. Three different versions of such CFA-aware features were introduced, while their detection performance was examined. Five different demosaicking algorithms were used for two steganographic methods LSB matching and WOW. It was found that the accuracy of detection and the boost from CFA is greatly influenced by the demosaicking algorithm and, generally, on the RAW-to-RGB converter. The detection of images which were processed using bilinear and VNG demosaicking in ufaw and Adobe Lightroom was enabled by richification. Unlike for Lightroom images, detection is improved by the CFA awareness, and this is specifically important for small payloads for both WOW and LSB matching.

The unique features which enable the improvement of the stegos from cover in an increasing or decreasing manner, are extracted using the Relative Auto-Decorrelation (RAD) method of feature extraction [8]. In order to obtain rapid detection accuracy, while im-

proving the results of steganalysis, the right features like Local Entropies Sum (LES) and Clouds Min Sum (CMS) must be acquired. A procedure of smart partitioning is used in processing the common parts; this procedure is referred to as two-dimensional Decorrelation of the received images (2D). The quadratic estimator, which is capable of estimating the embedding rate, enables the identification of secret message. The thresholds which are obtained from RAC, LES and CMS are used in modifying the estimation rate [64].

A method which calculates the differences among pairs of pixels was proposed by Chen et al. [65], who also proved that stego noises can be levelled by the histogram of difference values. The difference histogram characteristic function (DHCF) as well as the moment of DHCFs (DHCFM) were calculated and used as unique features. Calibration of features was carried out by reducing the effect of image content on them, while an SVM classifier was trained using the calibrated features. The training and test set used were BOSSBase and NRCS, while the rate of embedding was 0.25bpp. Results of the experiment showed that the detection of stego messages concealed by LSB matching, is enabled by the DHCFMs calculated with nonadjacent pixels.

In this paper, the researchers proposed a combination of the LTP and pi-LBP features for the purpose of extracting important dissimilarities between the stego and cover images. A description of the dissimilarities from different angles is given using the proposed pi-LTP based on a variety of flexible parameters. The optimal subspace of pi-LTP features is further selected using a robust incremental algorithm so as to obtain a good balance between the feature dimensionality and discrimination. Results of experiments shows performed on the BOSSbase 1.01 database demonstrates that the superiority of the proposed method [66].

Through the use of Fisher Criterion and ANOVA techniques, a reduced dimensional merged feature was proposed by Desai et al. [67] for universal image steganalysis. The extraction of features was made from binary similarity patterns and wavelet which were also extracted from DCT of an image. These features were combined to produce a combined feature set. In order to evaluate the combined feature vector score, Fisher criterion and ANOVA test were used. Afterwards only features that were sensitive in both feature selection methods were selected. The SVM classifier with RBF kernel was trained using the reduced 15-dimensional feature vector. The proposed algorithm was tested in comparison to other methods of steganography at different rates of embedding.

In [68], a method of detection for LSB flipping embedding method was proposed by Chaeikar et al. The contribution of their work can be seen from five different perspectives. First, it is a method that can be used to analyze the correlativity of pixel in pixel similarity weight (PSW). Second, deviating pixels, which are detected statistically are eliminated. Third, from the statistically detected pixels, ranking order was done, and effect of those pixels were determined. Then an analysis of the classes of pixel was done; the classes given to the pixels include edgy, flat and smooth class.

The modified version of the RS-steganalysis for BMP stego-images is offered based on applying the method to different size groups of pixels. The tribological statistics of stego-program, which is described as the traces of LSB embedding are accumulated. Based on the known program steganalytic attack, more precise results are obtained using the modified version [69].

This study argues that high detection accuracy can be achieved by the steganalysis, if more attention is paid to the parts of the cover image that has higher probability of being used for embedding. In the proposed feature set, GLCM statistical texture features, correlation between the left and right half-bytes, Entropy of the right half-bytes, coefficient of variation of right half-bytes, and the absolute difference between successive right half-bytes are included. In the experiment, 10,000 of each clean, 2LSB stego and 4LSB stego images were analyzed using the SVM classifier [70].

In [71] Sandoval et al., the probability density function (PDF) of adjacent pixels and co-occurrence of the image was used in selecting 12 most significant features. The use of this feature vector was employed in training an SVM to be able to differentiate stego images from natural ones. Two image data sets were used in evaluating the proposed steganalysis scheme; the two data sets include BOWS and UCID [67] under four different embedding rates or payloads

Thach et.al [72] noted that the previous work needs to identify the modified pixels or residuals as an artifact of the process of embedding. Therefore, in their paper they provide good results that deals with the shortcomings. Results demonstrate that sufficient data is contained in the expected mean residuals to order logically the located payload, as long as there is variation in the size of payload in each stego image.

In this paper, the researchers propose a modified version of the EPoV analysis for the detection of 2LSB replacement in still images. The standard deviation of the frequency of occurrences in EPoVs is used here for the purpose of estimating the length of the hidden message, which instead of being a discrete classifier, becomes a probabilistic classifier. A set of 3000 never-compressed images12 were evaluated after they were converted to grey-scale with streams of quasi random binary values as a secret message; to make it very close to the encrypted version11. A comparison of the results is made with results of the method which was proposed by Niu et al.9. The results revealed that detection is more accurate in for low embedding rates in the proposed method [73].

A method of unsupervised steganalysis, which combines supervised classification and artificial training sets, was provided by Lerch-Hostalot et al. [74]. In this method, it is assumed that the approximate bit rate and embedded algorithm used by the steganographer are known. In order to produce stego images with varying embedding rates (0.10bpp, 0.20bpp, 0.25bpp and 0.40bpp), BOSSBase image set was used. The model has been tested on three steganography methods, with detailed comparative experiments performed.

Table 2: Summarized Presentation of LSB Substitution LSB Matching Steganalysis Methods

S. No	Method - feature	Accuracy - Comments
35	Chi-squared detects of POVs	• Different tests depending on steganography and size of embedding message. • Work with any steganography, which has a fixed set of pairs of values. • Reliably detects when the message placement has known also when sequentially embedding the message . • Later was generalized to detect randomly scattered messages [42–43] • Different tests of threshold and error probability for several different test message sizes.
36	• Statistical analysis features • Raw Quick Pairs method (RQP)	• The only color image is used (Limitation). • Reliably Work well as long as the number of unique colors in the cover image is less than 30% of the number of pixels. • Reliably to show the existing message only, cannot estimate the hiding message (Limitation). • Not work if the image is scaled down (Limitation).
37	RS steganalysis	• Higher accuracy detection than Chi-Square [35]. • Detect the randomly scattered message with estimate message length.

		• Originally aimed at color bitmaps. • Many tests and results depending on the initial bias. • The reliability is increased using different masks ،this improved the gray scale image also. • More reliable than Chi-square algorithm [35]. • Slightly better than [44] for a message that has shorter than 80% of capacity. • The message that requires less than 0.005 bpp is undetectable in this method. • The estimate can be extremely accurate (often within 1%). • Various tests and results depending on embedding message length • This method has improved in method [39].
38	Finite state Machine	• It was inspired by the work of [37]. • Offers an analytical explanation of an observation made in [37] and sheds light on why [37] achieved such remarkable accuracy and efficacy. • Detect LSB at continuous tone natural images. • Different tests of threshold and error probability for several different test message sizes. • The only color image is used (Limitation).
39	Finite state Machine	• Higher accuracy detection than POVs [35]. • Reliably Work well as long as the number of unique colors in the cover image is less than 30% of the number of pixels. • Reliably to show the existing message only, cannot estimate the hiding message (Limitation). • Not work if the image is scaled down (Limitation). • Higher accuracy detection than Chi-Square [35].
40	- Finite state Machine - FPA	• Different tests and results. • Used both color and gray scale images for test. • Focused on the countermeasures against LSB. • Detected the LSB of continuous signals. • Estimate accurately when the embedding rate is larger than 3%. • Different tests and results depending on embedding message length.
41	Pairs Analysis	• Inspired by work of RS [37] and it is a generalized case of [25,26]. • Reliably to detects the message even when the message is very short relative to the size of an image. • Outperforms the Chi-square attack [35] • Pairs Analysis is slightly worse than RS steganalysis [37]. • Different tests and results. • Outperformance than the chi-square [35].
42	Gradient Energy-Flipping Rate Detection (GEFR)	• Originally aimed at palette image. • There is no theoretical reason why this algorithm should not work with the grayscale image, this algorithm was improved to works well on grayscale images [81]. • It is not reliable for an image, which it does not hold the assumption of equal homogeneity. • Improved by excluding non-adjacent pixels from the homogeneity calculation. • Various tests and results depending on embedding rate.
43	Image Histogram	• Where embedding rate is >0.05 bpp, so it can rely on detects the secret message. • To estimate the message length, the GE curve is help with the straight line. • Can work with both gray scale & color images. • Works well both for sequential or random LSB substitution. • Better performance and computation speed than RS analysis [37].
44	Laplace distribution	• Various tests and results depending on embedding rate 0%, 62.5%, 75%, 87%, and 100. • Low computational complexity and fast computational speed.
45	Ker's sequential WS variant	• Outperforms on known WS method with a very small payload capacity. • Works only with the LSB replacement (Limitation). • Focus on adaptive embedding (Limitation). • Confirms the result qualitatively and quantitatively
46	• kSVR in SPAM • Rich models • kSVR in CCPEV • FLD as a classifier • Online ensemble average perceptive	• Outperforms prior quantitative steganalysis across all tested algorithms. • Features extraction of both domains are available on http://www.bing-hamton.edu. • New base linear can capable of locating estimation, it is variant of a regression tree [82].
47	• Principal Component Analysis (PCA) • SVM as a classifier. • Noise estimation of seventh and eighth levels • 5 features as RS method output.	• Grayscale images are used only. • Different datasets are used. • Various tests and results depending on embedding rate. • This method comes from combining two methods [37, 80].
48	• Hypothetical testing theory • GLRT • Additive White Gaussian Noise (AWGN) • Clipped phenomenon caused denoising. • Maximum Likelihood (ML).	• Still ensures a high detection performance while other detectors cannot tolerate the impact of clipping phenomenon and fail in practice. • Various tests for different embedding rates. • Used to exploit the denoising method used in [83]. • Comparison the revised version of WS [85] and AUMP detector [84].

49	• PVD histogram. • SVM with the Gaussian kernel as a classifier.	• Different tests and results depending on the size of embedding message. (10%, 20%, 30%, 40%, 50%, and 75% of the maximum size of embedding capacity using the EA-LSBMR steganography). • Effective detected in the EA-LSBMR steganography. • High performance when the payload is low. • High accuracy as compared to LLPDF [86] and (SPAM) [87]. • Used two datasets to evaluate. • Different tests and results depending on embedding rate. • Outperform all variant of WS analysis in both compressed and uncompressed images. • Low level of noise such as decompressed JPEGS or low pass filtered image.
50	• Parity- Aware features . • The co-occurrences neighboring feature. • Vector feature. • Simple feature of selection algorithm (FFS, wrapper).	• The feature of parity aware residual especially effective for steganalysis uncompressed images. • Easy to implement and it is used in about 70% of available steganography software on the internet. • Feature-based detectors with parity-aware features can significantly outperform all structural detectors as well as variants of WS analysis in both decompressed JPEG images and in uncompressed images. • Outperformed using as few as three co-occurrence bins in decompressed JPEGs and thirty bins for uncompressed images.
51	• Likelihood Ratio Test (LRT). • Asymptotically Uniformly Most Powerful (AUMP) • Generalized Likelihood Ratio Test (GLRT).	• Numerical results and comparisons with prior art detectors highlight the relevance of the proposed methodology • Embedding rate is lower than 25%.
52	• 16 dimensional feature vector. • FISHER linear classification. • Gray level Co-occurrence Matrix (GLCM).	• The detection accuracy is increased above 20% compared with the improved HCF feature‘ • Increased about 10% compared with the feature without using estimation image and is increased about 3% compared with WAM feature.
53	• Most Powerful (MP) • Constant False Alarm Rate (CFAR) • Generalized Likelihood Ratio Test (GLRT) • central limit theorem (CLT) • autoregressive (AR)	• Tests achieve a better detection power . • Can be applied on digital images and audio signals contrary to the prior-art detectors which mainly focus on digital images.
54	• Asymptotically Uniformly Most Powerful (AUMP). • Likelihood Ratio Test (LRT). • Multivariate Gaussian distribution (MGD).	• Various tests and results • Perform well on large databases. • Low accuracy detection than the method in [32]. • Higher detection accuracy than 2D-HCF detector [17], ALE [19] and [88].
55	• Compressive sensing algorithm (BM3D).	• Compressive sensing algorithms have been used, especially the BM3D algorithm. • Very fast and can process many pictures and detect stego content in seconds without any previous training. • High accuracy more than the previous work [54].
56	• likelihood ratio test • 2 Pixel Selection (PS). • Multivariate Gaussian distribution (MGD).	• The curve of our method doesn’t perform that well as [54] at first. • Better after the intersection while the AUC is obviously improved. • The AUC is improved from 0.7372 to 0.8059 and with the PS, the result performs even better, and the AUC is increased to 0.8595.
57	• Features extracted by three-dimensional co-occurrences of residuals computed from all three-color channels. • SRMQ1, CRMQ1, and their union SCRMQ1.	• Various tests for different embedding rates (0.05–0.5bpp) with average detection error as metric • Focuses on detection of both non-adaptive LSB matching and modern content-adaptive steganography in true color images in raster formats that were not previously subjected to JPEG compression. • Outperformance more than the 2D HCF COM [17], ALE [19] and SPAM [89].
58	• Calculation of the center of mass (COM) of the characteristic function of difference histogram (DHCF). • SVM as a classifier.	• The improvement achieved by the features extracted from nonadjacent pixel pairs are not remarkable. • Various test on different embedding rate (0.10–1.0bpp). • Two different datasets have been used for evaluation. • Minimized classification error as metric.
59	• Local binary patterns (LBPs). • SVM as a classifier.	• Detection ability is limited due to lacking the correlation of multi-scale pixels • Various tests and results on different embedding rate.
60	• Histogram Characteristic Function Center of Mass (HCF-CoM) • Probability Density Function (PDF)	•
61	• - singular value decomposition (SVD) • - SVM as a classifier.	• Various tests and results on different embedding rate. • Outperformance more than the different methods (LogSV, NJ, NP & SPAM). • Wm is selected from 5 to 8, to achieve a higher accuracy with acceptable computational cost.
63	• Color-rich model (CRM) • Bayer color filter array (CFA) • FLD as a classifier.	• Different tests depending on steganography tools and size of embedding message. • Higher performance accuracy compared with other methods.
64	• Relative auto-decorrelation (RAD) • Local-Entropies-Sum (LES)	• Different tests and results depending on embedding rate and comparison to prior methods.

	- Clouds-Min-Sum (CMS) - Cross-decorrelation (CD) - SVM and multi-order polynomial kernel.	• Different standard databases of images were used.
65	• Difference histogram characteristic function (DHCF) • Moment of DHCFs (DHCFM) is used as discriminative features. Features were calibrated by decreasing the influence of image content on them. • SVM classifier was trained	• Various test for embedding rate 0.25bpp. Results in paper figures. • High performance comparing with different methods.
66	• Local ternary pattern based on path integral (pi-LTP) • Local binary pattern (LBP) • SVM as a classifier.	• Better performance than the state-of-the-art steganalysis methods. • The comparison of AUC of LTP with the different embedding ratio.
67	• ANOVA • BSM • Fisher Criterion. • SVM and RBF kernel as classifiers.	• Achieve overall 97% detection accuracy. • Higher performance when compared to existing methods [91,92,93,94]. • ANOVA feature is used as reduction feature. • The average time is slower than other methods. • Different tests depending on steganography tools and size of embedding message.
68	• Pixel Similarity Weight (PSW) • statistically detected color correlativity regression • SVM is used as a classifier.	• It detects and then involves only suspicious pixels in the steganalysis process. • It defines the influence of suspicious pixels in the steganalysis process based on statistically detected color correlativity regression levels‘ • Analyzing suspicious pixels in two ways – first unified and then according to pixel class – thereby enhancing the sensitivity of the SVM steganalyzer • Outstanding efficiency rate of 98.049% in detecting 0.25bpp stego images with only a single dimension analysis. • High computational complexity (Limitation).
69	• Modification of RS-steganalysis and known program steganalysis attack.	• Various tests and results depending on different LSB embedding algorithm and different embedding rates (10%, 30%, 50%, 70% and 90%). • More reliable than different embedding algorithms such as CryptArkan, StegoMagic, S-Tools and the Third Eye. • Detected accuracy is 99.41% for 4LSB and 99.02% for 4LSB. • Focus on LSB only (Limitation).
70	• GLCM • SVM is used as a classifier.	• Greyscale images are chosen as the cover image (Limitation). • RHB is used, to get an attention to the area which is most embedding accuracy regardless. • Focus on part of the cover image, where it is more likely to be used for embedding.
71	• 12 relevant features based on the probability density function (PDF) of the difference between adjacent pixels and the co-occurrence matrix of the image . • SVM as classifier	• Various tests for different embedding rates (100%, 75%, 50%, 25%). • 87.2% detection accuracy. • Better discriminate performance than previous methods that require a larger amount of feature elements, such as 27, 35 and 225 features for their discriminations.
72	• GLCM • forensics features • LS-SVM is used as a classifier. • Extended Pairs of Values (EPoV)	• Different tests and results. • Two methods are used for cost training and testing comparison. • Outperforms the weighted stego method (WS2) for low embedding rates (less than 50%). • The detection rate is higher than [90]. • The Image feature is sensitive (changeful) for embedding. • Accurately estimate the length of the hidden message for any embedding rate.
73	Unsupervised steganalysis method combined with artificial training sets and supervised classification.	• Different tests for different embedding rates (0.1bpp, 0.2bpp, 0.25bpp, 0.4bpp) for three different steganographic algorithms. • Outperforms state of the art methods while running on the same order of the features. • Different data set and different size has been used to obtain a better performance.
74	• Local Filter Pattern (LFP) • Discretized-All Condensed Nearest Neighbour (D-All-CNN) • Co-occurrence matrix • Condensed Nearest Neighbour (CNN) • Greedy Randomised Adaptive Search Procedure (GRASP) • Recursive Feature Elimination (RFE)	• Different tests and results depending on embedding rate and comparison to prior methods • Efficiently used in improving the existing multi-model steganalysis features • High accuracy for both traditional non-adaptive and content adaptive.

Results showed that better classification accuracy is achieved by the proposed method than that of the conventional supervised steganalysis such as Esemble Classifiers, Rich Models etc.

In the proposed trilevel optimization, the feature model which can be formed using concatenation is optimized. Here, Greedy Randomised Adaptive Search Procedure (GRASP) is used in combining feature models. GRASP is described as a meta heuristics optimisation algorithm based on semi-greedy approach which is discussed in Section 4.1. In the second level the best instances that can improve the solution for optimal concatenated model is chosen. The paper proposes a new, iterative wrapper which is based on condensed incremental instance selection method known as AllCNN. Arnaiz-Gonzalez et al. [27] recently recommended the application of instance selection for general regression tasks by making the target value discrete. Therefore, Section 4.2 contains a detailed description of the proposed simple discretized AllCNN (D-AllCNN) instance method of selection. This method enables the selection of training samples within a quantitative steganalysis environment [75].

Table 2 shows the summarization of the main method, features, accuracy of the detection and important comments of the techniques.

9. Evaluation criteria of steganalysis

Steganalysis specifically aims at determining if a suspected medium is embedded with secret data. This means that, its main purpose is to ascertain if the testing medium belongs to the stego or cover class. Four possible situations could emerge if a particular kind of stageanalysis method is used to analyze a suspicious medium, and they are as follows:

- True positive (TP), which implies the classification of a medium as a stego medium is correct.
- False negative (FN), meaning that the classification of stego medium as a cover is wrong.
- True negative (TN), meaning that the classification of a cover medium as a cover is correct.
- False positive (FP), implying that classifying a cover medium as a stego is wrong.

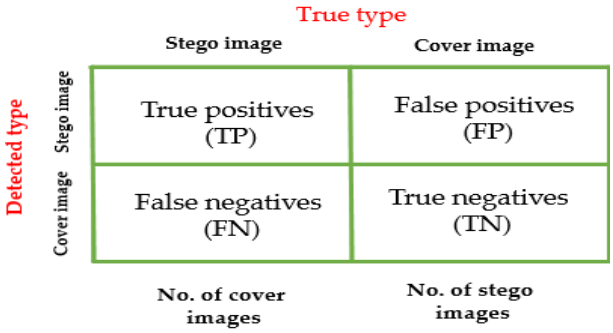

Fig. 7: Steganalysis Evaluation Confusion Matrix.

9.1. Confusion matrix

A 2 x 2 confusion matrix [8] can be developed when a steganalysis method is being applied on a testing set, which may be made up of stego and cover media as demonstrated in Figure 7. The constructed 2 x 2 confusion matrix can be constructed in a manner that dispositions of the instances in the set are represented in it. Some metrics of evaluation can be defined based on this constructed.

9.2. Receiver operating characteristic (ROC) curve

A ROC curve can be used in visualizing the performance of a steganalysis classifier [8]. On the ROC curve, the vertical axis contains the plotting of the true positive rate, while the false positive rate is plotted on the horizontal axis as illustrated in figure 8. The performance of the steganalysis method will be better if the area under the ROC curve (AUC) is larger. For instance, Figure 8 shows that the performance of ROC curve C is better than B, while B is better than A. The figure 9 is showing the example of ROC curve.

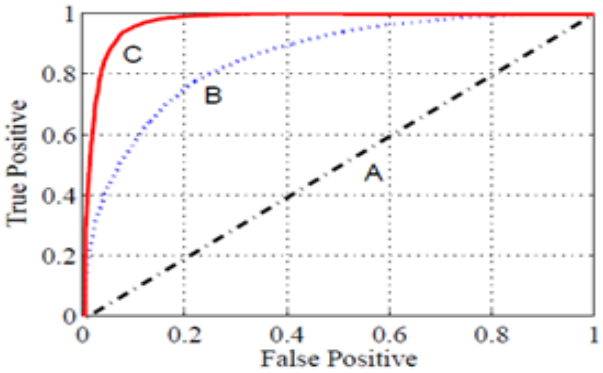

Fig. 8: Steganalysis Evaluation ROC Curve [75].

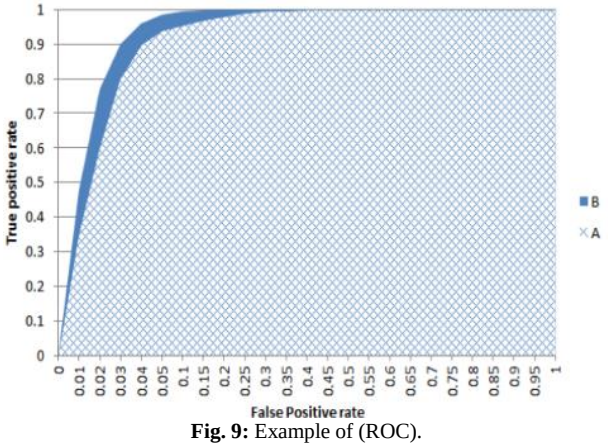

Fig. 9: Example of (ROC).

In Table 3, we summarize the methods along with dataset and number of images, regarding the dataset the authors utilized through their papers.

Table 3: Summarized the Methods, Dataset and Number of Images

S. No	year	Detected algorithm	Database	No. of image
35	2000	Steganos EzStego LSB substitution S-Tools	Unknown database images	5
36	2000	LSB substitution	Unknown database images,	300
37	2001	Steganos, S-Tools, Hide4PGP	Unknown database images	3
38	2002	LSB substitution	Unknown database images	29
39	2002	LSB substitution	Unknown database images	29
40	2003	LSB substitution	Unknown database images	29
41	2003	LSB substitution, EzStego	Color GIF images obtained using four different digital cameras, originally stored as high-quality JPEG images	180
42	2003	LSB substitution	Unknown database images	4
43	2003	LSB matching	USC-SIP1	5
44	2010	LSB matching	NRCS database UCID database	3185
45	2012	LSB replacement	BOSSbase database	10000
46	2013	LSBR, HUGO, nsF5, BCHOPT and MME3	BOSSbase ver. 0.92 database	10000
47	2013	LSB substitution	NRCS database USC database Corel database	15108

48	2013	LSB substitution	USID database DB2 database Dresden database BOSSbase database	5000
49	2013	EA-LSBMR	NRCS database UCID database	2000
50	2012	LSB matching	BOSSbase database	10000
51	2013	LSB matching	BOSSbase database	10000
52	2013	LSB matching	No reference by the authors,	485
53	2013	LSB matching	BOSSbase database	10000
54	2013	LSB matching	BOWS database BOSSBase v1.01 database	10000
55	2014	LSB substitution and DCT (Steghide)	No known database used	10
56	2015	LSB matching	BOSSBase v1.01 database	10000
57	2014	LSB matching and WOW.	BOSSBase v1.01 database	10000
58	2014	LSB matching	BOSSBase database NRCS database	12,644
59	2014	LSB matching	BOSSbase database	10000
60	2015	LSB Matching	UCID	15000
61	2015	LSB substitution, Stghide, PQ and F5	- UCID database	2024
63	2015	WOW and LSB Matching	BOSSbase 1.01 database	10000
64	2015	LSBM and LSBR	BOSSBase database COREL database NRCS database	10 000
65	2016	LSB Matching	BOSSBase database NRCS database	12335
66	2016	LSB Matching	BOSSbase 1.01 database	10000
67	2016	LSB, F5 and Outguess	BSDS500 database COREL database	1400
69	2017	CryptArkan, StegoMagic and S-Tools	Unknown data set	Unknown
73	2017	LSB substitution	BOSS database BOSSrank NRCS database ESO database Interactions database NOAA database Albion database Calphotos database	19000
72	2015	2LSB replacement	Dataset of Ref [62]	3000
70	2017	2LSB and 4LSB	BOSSbase1.01 database	15000
71	2017	LSB Matching	BOWS database UCID database	11228
68	2018	LSB substitution	University of (UW) image database	1333
74	2018	HUGO, WOW, SW LSBR, LSBM, LSBMR, LSBR2 and LSBR mod5	BOSSBase database	5000

10. Conclusion

The research in this paper is all about the background analysis and elaborate of statistical steganalysis in LSB substitution and LSB matching approach also provides a detailed reference of earlier steganalysis methods to state of the art for the digital images. The main contribution of the paper is to analyse the challenges and comparison of approved studies to evaluate the methods and open interesting directions that provide a better way for an effective steganalysis approach. The performance and evaluation analysis of the techniques discussed in the paper have also been given, by using metrics such as the detection rate, the error rate and ROC curves in specific embedding rates. Finally, in Table 4 we summarize the research, regarding the dataset the authors utilized through their papers. Their full names of datasets along with the download links can be get in the appendix.

Acknowledgement

This research was supported by the School of Computing, University Technology Malaysia (UTM), Johor Bahru, Malaysia. The authors would like to thank the reviewers for their insightful comments and helpful suggestions for a better research.

Appendix - Dataset Links

.	Dataset links	
1	Philip Greenspun	http://philip.greenspun.com/
2	Noname	http://www.petitcolas.net/fabien/watermarking/benchmark/image_database.html
3	CBlR Image	http://imagedatabase.cs.washington.edu/groundtruth/
4	BOOSBase	http://agents.fel.cvut.cz/boss/index.php?mode=VIEW%26tmpl=materials
5	Corel	https://www.corel.com/en/
7	UCID	http://jasoncantarella.com/downloads/
8	Noname	https://www.cs.nmt.edu/~IA/steganalysis.html
9	BSDS	https://www2.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/fg/
10	BSDS500	https://www2.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/segbench/
11	KODAK	http://ftp//ftp.kodak.com/www/images/pcd/
12	USC-SIP1 Image	http://sipi.usc.edu/database/
13	NRCS	https://photogallery.sc.egov.usda.gov/res/sites/photogallery/
14	Noname	http://old.vision.ece.ucsb.edu/~sullivak
15	BOWS 2	http://bows2.ec-lille.fr/
16	INRIA	http://lear.inrialpes.fr/~jegou/data.php
17	ImageNet	http://www.image-net.org/
18	Raise	http://mmlab.science.unitn.it/RAISE/
19	Washington University (um)	http://imagedatabase.cs.washington.edu/

References

[1] Srivastava, S., Thakral, P., Bansal, V., & Shandil, V. (2018). A Novel Image Steganography and Steganalysis Technique Based on Pattern Searching. In Optical and Wireless Technologies (pp. 531-537). Springer, Singapore. https://doi.org/10.1007/978-981-10-7395-3_59.

[2] Boroumand, M., & Fridrich, J. (2018). Applications of explicit non-linear feature maps in steganalysis. IEEE Transactions on Information Forensics and Security, 13(4), 823-833. https://doi.org/10.1109/TIFS.2017.2766580.

[3] Karampidis, K., Kavallieratou, E., & Papadourakis, G. (2018). A review of image steganalysis techniques for digital forensics. Journal of information security and applications, 40, 217-235. https://doi.org/10.1016/j.jisa.2018.04.005.

[4] MAHDI HASHIM, M. O. H. A. M. M. E. D., RAHIM, M., & SHAFRY, M. (2017). IMAGE STEGANOGRAPHY BASED ON ODD/EVEN PIXELS DISTRIBUTION SCHEME AND TWO PARAMETERS RANDOM FUNCTION. Journal of Theoretical & Applied Information Technology, 95(22).

[5] Johnson, N. F., & Jajodia, S. (1998, April). Steganalysis of images created using current steganography software. In International Workshop on Information Hiding (pp. 273-289). Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-49380-8_19.

[6] Johnson, N. F., & Jajodia, S. (1998). Exploring steganography: Seeing the unseen. Computer, 31(2). https://doi.org/10.1109/MC.1998.4655281.

[7] Geetha, S., Sindhu, S. S. S., & Kamaraj, N. (2010). Passive steganalysis based on higher order image statistics of curvelet transform. International Journal of Automation and Computing, 7(4), 531-542. https://doi.org/10.1007/s11633-010-0537-1.

[8] Chandramouli, R., Kharrazi, M., & Memon, N. (2003, October). Image steganography and steganalysis: Concepts and practice. In International Workshop on Digital Watermarking (pp. 35-49). Springer, Berlin, Heidelberg.

[9] HASHIM, M., RAHIM, M., SHAFRY, M., & ALWAN, A. A. (2018). A REVIEW AND OPEN ISSUES OF MULTIFARIOUS IMAGE STEGANOGRAPHY TECHNIQUES IN SPATIAL DOMAIN. Journal of Theoretical & Applied Information Technology, 96(4).

[10] Katzenbeisser, S., & Petitcolas, F. (2000). Information hiding techniques for steganography and digital watermarking. Artech house.

[11] Nissar, A., & Mir, A. H. (2010). Classification of steganalysis techniques: A study. Digital Signal Processing, 20(6), 1758-1770. https://doi.org/10.1016/j.dsp.2010.02.003.

[12] Chanu, Y. J., Singh, K. M., & Tuithung, T. (2012). Image steganography and steganalysis: A survey. International Journal of Computer Applications, 52(2).

[13] Luo, X. Y., Wang, D. S., Wang, P., & Liu, F. L. (2008). A review on blind detection for image steganography. Signal Processing, 88(9), 2138-2157. https://doi.org/10.1016/j.sigpro.2008.03.016.

[14] Pal, P., & Dubey, S. (2016). Various JPEG image steganography techniques: a review. International Journal of Scientific & Engineering Research, 7(2), 417-421.

[15] Juarez-Sandoval, O., Cedillo-Hernandez, M., Sanchez-Perez, G., Toscano-Medina, K., Perez-Meana, H., & Nakano-Miyatake, M. (2017, April). Compact image steganalysis for LSB-matching steganography. In Biometrics and Forensics (IWBF), 2017 5th International Workshop on (pp. 1-6). IEEE.

[16] Mielikainen, J. (2006). LSB matching revisited. IEEE signal processing letters, 13(5), 285-287. https://doi.org/10.1109/LSP.2006.870357.

[17] Ker, A. D. (2005). Steganalysis of LSB matching in grayscale images. IEEE signal processing letters, 12(6), 441-444. https://doi.org/10.1109/LSP.2005.847889.

[18] Xi, L., Ping, X., & Zhang, T. (2010, July). Improved LSB matching steganography resisting histogram attacks. In Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference on (Vol. 1, pp. 203-206). IEEE.

[19] Zhang, J., Cox, I. J., & Doerr, G. (2007, October). Steganalysis for LSB matching in images with high-frequency noise. In Multimedia Signal Processing, 2007. MMSP 2007. IEEE 9th Workshop on (pp. 385-388). IEEE.

[20] Wayner, P. (2002). Disappearing cryptography information hiding: steganography & watermarking (2nd ed). Amsterdam: Morgan Kaufmann Publishers.

[21] Westfeld, A., & Pfitzmann, A. (2000). Attacks on steganographic systems breaking the steganographic utilities EzStego, Jsteg, Steganos, and S-Tools—and some lessons learned Lecture notes in computer science. vol. 1768. Berlin: Springer-Verlag.

[22] Zhang, T., & Ping, X. (2003, April). Reliable detection of LSB steganography based on the difference image histogram. In Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP'03) 2003 IEEE International Conference on (Vol. 3, pp. III-545). IEEE.

[23] Karampidis, K., Kavallieratou, E., & Papadourakis, G. (2018). A review of image steganalysis techniques for digital forensics. Journal of information security and applications, 40, 217-235. https://doi.org/10.1016/j.jisa.2018.04.005.

[24] Johnson, N. F., & Jajodia, S. (1998, April). Steganalysis of images created using current steganography software. In International Workshop on Information Hiding (pp. 273-289). Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-49380-8_19.

[25] Hawi, T. A., Qutayri, M. A., & Barada, H. (2004, November). Steganalysis attacks on stego-images using stego-signatures and statistical image properties. In TENCON 2004. 2004 IEEE Region 10 Conference (pp. 104-107). IEEE.

[26] Niimi, M., Eason, R. O., Noda, H., & Kawaguchi, E. (2001, August). Intensity histogram steganalysis in BPCS-steganography. In Security and Watermarking of Multimedia Contents III (Vol. 4314, pp. 555-565). International Society for Optics and Photonics. https://doi.org/10.1117/12.435440.

[27] Zhi, L., Fen, S. A., & Xian, Y. Y. (2003, September). A LSB steganography detection algorithm. In Personal, Indoor and Mobile Radio Communications, 2003. PIMRC 2003.14th IEEE Proceedings on (Vol. 3, pp. 2780-2783). IEEE.

[28] Chandramouli, R., & Subbalakshmi, K. P. (2003, May). Active steganalysis of spread spectrum image steganography. In Circuits and Systems, 2003. ISCAS'03. Proceedings of the 2003 International Symposium on (Vol. 3, pp. III-III). IEEE.

[29] Yu, X., Tan, T., & Wang, Y. (2004, December). Reliable detection of BPCS-steganography in natural images. In null (pp. 333-336). IEEE.

[30] Fridrich, J., Goljan, M., & Hogea, D. (2002, October). Steganalysis of JPEG images: Breaking the F5 algorithm. In International Workshop on Information Hiding (pp. 310-323). Springer, Berlin, Heidelberg.

[31] Jiang, M., Wong, E. K., Memon, N., & Wu, X. (2005, March). Steganalysis of halftone images. In Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP'05). IEEE International Conference on (Vol. 2, pp. ii-793). IEEE.

[32] Farid, H. (2002). Detecting hidden messages using higher-order statistical models. In Image Processing. 2002. Proceedings. 2002 International Conference on (Vol. 2, pp. II-II). IEEE. https://doi.org/10.1109/ICIP.2002.1040098.

[33] Avcibas, I., Memon, N., & Sankur, B. (2003). Steganalysis using image quality metrics. IEEE transactions on Image Processing, 12(2), 221-229. https://doi.org/10.1109/TIP.2002.807363.

[34] Harmsen, J. J., & Pearlman, W. A. (2003, June). Steganalysis of additive-noise modelable information hiding. In Security and Watermarking of Multimedia Contents V (Vol. 5020, pp. 131-143). International Society for Optics and Photonics. https://doi.org/10.1117/12.476813.

[35] Westfeld, A., & Pfitzmann, A. (1999, September). Attacks on steganographic systems. In International workshop on information hiding (pp. 61-76). Springer, Berlin, Heidelberg.

[36] Fridrich, J., & Long, M. (2000). Steganalysis of LSB encoding in color images. In Multimedia and Expo, 2000. ICME 2000. 2000 IEEE International Conference on (Vol. 3, pp. 1279-1282). IEEE. https://doi.org/10.1109/ICME.2000.871000.

[37] Fridrich, J., Goljan, M., & Du, R. (2001). Detecting LSB steganography in color, and gray-scale images. IEEE multimedia, 8(4), 22-28. https://doi.org/10.1109/93.959097

[38] Dumitrescu, S., Wu, X., & Memon, N. (2002, June). On steganalysis of random LSB embedding in continuous-tone images. In Image Processing. 2002. Proceedings. 2002 International Conference on (Vol. 3, pp. 641-644). IEEE.

[39] Dumitrescu, S., Wu, X., & Wang, Z. (2002, October). Detection of LSB steganography via sample pair analysis. In *International Workshop on Information Hiding* (pp. 355-372). Springer, Berlin, Heidelberg.

[40] Dumitrescu, S., & Wu, X. (2002). Steganalysis of LSB embedding in multimedia signals. In Multimedia and Expo, 2002. ICME'02. Proceedings. 2002 IEEE International Conference on (Vol. 1, pp. 581-584). IEEE.

[41] Fridrich, J., Goljan, M., & Soukal, D. (2003, June). Higher-order statistical steganalysis of palette images. In Security and Watermarking of Multimedia Contents V (Vol. 5020, pp. 178-191). International Society for Optics and Photonics. https://doi.org/10.1117/12.473140.

[42] Westfeld, A. (2002, October). Detecting low embedding rates. In International Workshop on Information Hiding (pp. 324-339). Springer, Berlin, Heidelberg.

[43] Chandramouli, R., Kharrazi, M., & Memon, N. (2003, October). Image steganography and steganalysis: Concepts and practice. In International Workshop on Digital Watermarking (pp. 35-49). Springer, Berlin, Heidelberg.

[44] Zhang, T., Li, W., Zhang, Y., Zheng, E., & Ping, X. (2010). Steganalysis of LSB matching based on statistical modeling of pixel difference distributions. Information Sciences, 180(23), 4685-4694. https://doi.org/10.1016/j.ins.2010.07.037.

[45] Schottle, P., Korff, S., & Bohme, R. (2012, December). Weighted stego-image steganalysis for naive content-adaptive embedding. In Information Forensics and Security (WIFS), 2012 IEEE International Workshop on (pp. 193-198). IEEE.

[46] Kodovský, J., & Fridrich, J. (2013, March). Quantitative steganalysis using rich models. In Media Watermarking, Security, and Forensics 2013 (Vol. 8665, p. 86650O). International Society for Optics and Photonics.

[47] Mirjavadi, S., Hamouda, A. M. S., Panahi, M. S., Jebeli, S. M., & Mousavi, M. (2013). A Combined Approach for Steganalysis embedded data stored in gray-scale images through LSB.

[48] Thai, T. H., Retraint, F., & Cogranne, R. (2014). Statistical detection of data hidden in least significant bits of clipped images. Signal Processing, 98, 263-274. https://doi.org/10.1016/j.sigpro.2013.11.027.

[49] Zhu, Z., Zhang, T., & Wan, B. (2013, June). A special detector for the edge adaptive image steganography based on LSB matching revisited. In Control and Automation (ICCA), 2013 10th IEEE International Conference on (pp. 1363-1366). IEEE.

[50] Fridrich, J., & Kodovský, J. (2012, May). Steganalysis of LSB replacement using parity-aware features. In International Workshop on Information Hiding (pp. 31-45). Springer, Berlin, Heidelberg.

[51] Cogranne, R., Thai, T. H., & Retraint, F. (2013, September). Asymptotically optimal detection of LSB matching data hiding. In Image Processing (ICIP), 2013 20th IEEE International Conference on (pp. 4437-4441). IEEE.

[52] Guo, Y. Q., Kong, X. W., Wang, B., & Xiao, Q. (2013). Steganalysis of LSB matching based on the sum features of average co-occurrence matrix using image estimation. In The International Workshop on Digital Forensics and Watermarking 2012 (pp. 34-43). Springer, Berlin, Heidelberg.

[53] Cogranne, R., & Retraint, F. (2013). Application of hypothesis testing theory for optimal detection of LSB matching data hiding. Signal Processing, 93(7), 1724-1737. https://doi.org/10.1016/j.sigpro.2013.01.014.

[54] Cogranne, R., & Retraint, F. (2013). An asymptotically uniformly most powerful test for LSB matching detection. IEEE transactions on information forensics and security, 8(3), 464-476. https://doi.org/10.1109/TIFS.2013.2238232.

[55] Patsakis, C., & Aroukatos, N. (2014). LSB and DCT steganographic detection using compressive sensing. Journal of Information Hiding and Multimedia Signal Processing, 5(1), 20-32.

[56] Yang, G., Li, X., Li, B., & Guo, Z. (2015, December). A new detector of LSB matching steganography based on likelihood ratio test for multivariate Gaussian covers. In Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2015 Asia-Pacific (pp. 757-760). IEEE.

[57] Goljan, M., Fridrich, J., & Cogranne, R. (2014, December). Rich model for steganalysis of color images. In Information Forensics and Security (WIFS), 2014 IEEE International Workshop on (pp. 185-190). IEEE.

[58] Xia, Z., Wang, X., Sun, X., Liu, Q., & Xiong, N. (2016). Steganalysis of LSB matching using differences between nonadjacent pixels. Multimedia Tools and Applications, 75(4), 1947-1962. https://doi.org/10.1007/s11042-014-2381-8.

[59] Gui, X., Li, X., & Yang, B. (2014, August). Steganalysis of LSB matching based on local binary patterns. In Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2014 Tenth International Conference on (pp. 475-480). IEEE.

[60] Olguin-Garcia, H. J., Juarez-Sandoval, O. U., Nakano-Miyatake, M., & Perez-Meana, H. (2015, January). Color image steganalysis method for LSB matching. In Proceedings of the International Conference on Security and Management (SAM) (p. 309). The Steering Committee of the World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp).

[61] Nouri, R., & Mansouri, A. (2015, November). Blind image steganalysis based on reciprocal singular value curve. In Machine Vision and Image Processing (MVIP), 2015 ninth Iranian Conference on (pp. 124-127). IEEE.

[62] Dataset http://www.shsu.edu/~qxl005/New/Downloads/index.html

[63] Goljan, M., & Fridrich, J. (2015, March). CFA-aware features for steganalysis of color images. In Media Watermarking, Security, and Forensics 2015 (Vol. 9409, p. 94090V). International Society for Optics and Photonics.

[64] Farhat, F., & Ghaemmaghami, S. (2014). Towards blind detection of low-rate spatial embedding in image steganalysis. IET Image Processing, 9(1), 31-42. https://doi.org/10.1049/iet-ipr.2013.0877.

[65] Chen, X., GAO, G., Liu, D., & Xia, Z. (2016). Steganalysis of LSB matching using characteristic function moment of pixel differences. China Communications, 13(7), 66-73. https://doi.org/10.1109/CC.2016.7559077.

[66] Lin, Q., Liu, J., & Guo, Z. (2016, September). Local ternary pattern based on path integral for steganalysis. In Image Processing (ICIP), 2016 IEEE International Conference on (pp. 2737-2741). IEEE.

[67] Desai, M. B., Patel, S. V., & Prajapati, B. (2016). ANOVA and Fisher Criterion based feature selection for lower dimensional universal image steganalysis. International Journal of Image Processing (IJIP), 10(3), 145-160.

[68] Chaeikar, S. S., Zamani, M., Manaf, A. B. A., & Zeki, A. M. (2018). PSW statistical LSB image steganalysis. Multimedia Tools and Applications, 77(1), 805-835. https://doi.org/10.1007/s11042-016-4273-6.

[69] Solodukha, R. A., & Atlasov, I. V. (2017, September). Modification of RS-steganalysis to attacks based on known stego-program. In Computer Technology and Applications (RPC), 2017 Second Russia and Pacific Conference on (pp. 176-179). IEEE.

[70] Al-Jarrah, M. M., Al-Taie, Z. H., & Aboarqoub, A. (2017, July). Steganalysis Using LSB-Focused Statistical Features. In Proceedings of the International Conference on Future Networks and Distributed Systems (p. 41). ACM. https://doi.org/10.1145/3102304.3109814.

[71] Juarez-Sandoval, O., Cedillo-Hernandez, M., Sanchez-Perez, G., Toscano-Medina, K., Perez-Meana, H., & Nakano-Miyatake, M. (2017, April). Compact image steganalysis for LSB-matching steganography. In Biometrics and Forensics (IWBF), 2017 5th International Workshop on (pp. 1-6). IEEE.

[72] Khalind, O., & Aziz, B. (2015, May). A better detection of 2LSB steganography via standard deviation of the extended pairs of values. In *Mobile Multimedia/Image Processing, Security, and Applications 2015* (Vol. 9497, p. 94970E). International Society for Optics and Photonics.

[73] Lerch-Hostalot, D., & Megías, D. (2016). Unsupervised steganalysis based on artificial training sets. *Engineering Applications of Artificial Intelligence*, *50*, 45-59. https://doi.org/10.1016/j.engappai.2015.12.013.

[74] Veena, S. T., & Arivazhagan, S. (2018). Quantitative steganalysis of spatial LSB based stego images using reduced instances and features. *Pattern Recognition Letters*, *105*, 39-49. https://doi.org/10.1016/j.patrec.2017.08.01.

[75] Li, B., He, J., Huang, J., & Shi, Y. Q. (2011). A survey on image steganography and steganalysis. *Journal of Information Hiding and Multimedia Signal Processing*, *2*(2), 142-172.

[76] Fridrich, J., & Kodovsky, J. (2012). Rich models for steganalysis of digital images. IEEE Transactions on Information Forensics and Security, 7(3), 868-882. https://doi.org/10.1109/TIFS.2012.2190402.

[77] Kodovský, J., & Fridrich, J. (2012, February). Steganalysis of JPEG images using rich models. In Media Watermarking, Security, and Forensics 2012 (Vol. 8303, p. 83030A). International Society for Optics and Photonics.

[78] Pevný, T., Fridrich, J. J., & Ker, A. D. (2012). From blind to quantitative Steganalysis. *IEEE Trans. Information Forensics and Security*, *7*(2), 445-454. https://doi.org/10.1109/TIFS.2011.2175918.

[79] Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189-1232. https://doi.org/10.1214/aos/1013203451.

[80] Diyanat, A., Farhat, F., & Ghaemmaghami, S. (2011, November). Image steganalysis based on SVD and noise estimation: Improve sensitivity to spatial LSB embedding families. In TENCON 2011-2011 IEEE Region 10 Conference (pp. 1266-1270). IEEE.

[81] Ker, A. D. (2004, June). Quantitative evaluation of pairs and RS steganalysis. In Security, Steganography, and Watermarking of Multimedia Contents VI (Vol. 5306, pp. 83-98). International Society for Optics and Photonics. https://doi.org/10.1117/12.526720.

[82] L. Breiman, J. Friedman, R. Olshen, and C. Stone. Classification and Regression Trees. Wadsworth and Brooks, Monterey, CA, 1984.

[83] Foi, A. (2009). Clipped noisy images: Heteroskedastic modeling and practical denoising. *Signal Processing*, *89*(12), 2609-2629. https://doi.org/10.1016/j.sigpro.2009.04.035.

[84] L. Fillatre. Adaptive steganalysis of least significant bit replacement in grayscale natural images. Signal Processing, IEEE Transactions

on, 60(2):556 – 569, February 2012. https://doi.org/10.1109/TSP.2011.2174231.

[85] Ker, A. D., & Böhme, R. (2008, March). Revisiting weighted stego-image steganalysis. In *Security, Forensics, Steganography, and Watermarking of Multimedia Contents X* (Vol. 6819, p. 681905). International Society for Optics and Photonics. https://doi.org/10.1117/12.766820.

[86] Li, B., Huang, J., & Shi, Y. Q. (2008, March). Textural features based universal steganalysis. In *Security, Forensics, Steganography, and Watermarking of Multimedia Contents X* (Vol. 6819, p. 681912). International Society for Optics and Photonics. https://doi.org/10.1117/12.765817.

[87] Pevny, T., Bas, P., & Fridrich, J. (2010). Steganalysis by subtractive pixel adjacency matrix. *IEEE Transactions on information Forensics and Security*, *5*(2), 215-224. https://doi.org/10.1109/TIFS.2010.2045842.

[88] Cogranne, R., Zitzmann, C., Retraint, F., Nikiforov, I., Fillatre, L., & Cornu, P. (2012, May). Statistical detection of LSB matching using hypothesis testing theory. In *International Workshop on Information Hiding* (pp. 46-62). Springer, Berlin, Heidelberg.

[89] Pevny, T., Bas, P., & Fridrich, J. (2010). Steganalysis by subtractive pixel adjacency matrix. *IEEE Transactions on information Forensics and Security*, *5*(2), 215-224. https://doi.org/10.1109/TIFS.2010.2045842.

[90] Yu, X., & Babaguchi, N. (2008, June). Weighted stego-image based steganalysis in multiple least significant bits. In *Multimedia and Expo, 2008 IEEE International Conference on* (pp. 265-268). IEEE.

[91] Lyu, S., & Farid, H. (2002, October). Detecting hidden messages using higher-order statistics and support vector machines. In *International Workshop on Information Hiding* (pp. 340-354). Springer, Berlin, Heidelberg.

[92] Xuan, G., Shi, Y. Q., GAO, J., Zou, D., Yang, C., Zhang, Z., & Chen, W. (2005, June). Steganalysis based on multiple features formed by statistical moments of wavelet characteristic functions. In *International Workshop on Information Hiding* (pp. 262-277). Springer, Berlin, Heidelberg. https://doi.org/10.1007/11558859_20.

[93] Shi, Y. Q., Xuan, G., Yang, C., GAO, J., Zhang, Z., Chai, P. ... & Chen, W. (2005, April). Effective steganalysis based on statistical moments of wavelet characteristic function. In *Information Technology: Coding and Computing, 2005. ITCC 2005. International Conference on* (Vol. 1, pp. 768-773). IEEE.

[94] Lin, J. Q., & Zhong, S. P. (2009, July). JPEG image steganalysis method based on binary similarity measures. In *Machine Learning and Cybernetics, 2009 International Conference on* (Vol. 4, pp. 2238-2243). IEEE.

[95] DOMAIN, W. T. I. S. (2018). A REVIEW AND OPEN ISSUES OF DIVERSE TEXT WATERMARKING TECHNIQUES IN SPATIAL DOMAIN. *Journal of Theoretical and Applied Information Technology*, *96*(17).