Acta Cybernetica **24** (2020) 593–612.

An Information Theoretic Image Steganalysis for LSB Steganography

Sonam Chhikara[a] and Rajeev Kumar[b]

Abstract

Steganography hides the data within a media file in a subtle way while steganalysis exposes steganography by using detection measures. Traditionally, steganalysis reveals steganography by targeting perceptible and statistical properties of the image for improving the security of steganography methods. In this work, we target LSB image steganography methods by using information theoretic metrics for steganalysis. Our technique works in two phases. First, the features of embedded image are extracted, and then this image is analyzed on the basis of entropy and joint entropy features corresponding to the original image. Second, from these extracted features of the image, we train SVM and ensemble classifiers. The classifiers discriminate cover image from the stego image. For evaluating the robustness of our method, several attacks over the stego images are applied. These attacked stego images are then analyzed for the detection reliability of our method. Original images and LSB embedded images of the dataset are analyzed by comparing information gain from entropy and joint entropy metrics. Results suggest that entropy of the stego images are more preserving than that of joint entropy. Experimental results show that before histogram attack, detection rate with entropy and joint entropy are 70% and 98%, respectively while after the attack entropy metric gives 30% detection rate and joint entropy gives 93% detection rate. Therefore, joint entropy proves to be a good steganalysis measure having fewer false alarms for across a range of hiding ratios.

Keywords: data embedding, steganography, information theory, steganalysis, classifier

1 Introduction

Steganography aims to support secret communication in an innocuous looking media file [1]. The core objective of steganography is to maximize secret information (embedding capacity) with least distortion in original cover media (imperceptibility). Cover media and secret data both can be audio, video, image, and a text

[a] Deceased on Dec. 16, 2019. This paper is dedicated to her memory.

[b] School of Computer and Systems Sciences, Jawaharlal Nehru University, India. E-mail: `RajeevKumar.cse@gmail.com`, ORCID: `https://orcid.org/0000-0003-0233-6563`.

DOI: 10.14232/actacyb.279174

file. Imperceptibility indicates that embedding should not be visually recognizable. However, due to the addition of secret information, steganography subtly degrades the embedding media quality. Higher embedding capacity results in more modifications in original media that affects imperceptibility. Therefore, to overcome this trade-off a good steganography technique focuses on both the requirements in a balanced way. Image steganography methods are generally classified into spatial and transform based techniques.

Spatial based steganography hides data directly in image pixels e.g., LSB embedding. LSB embedding is the simplest method for hiding the secret data by replacing the least significant bit of each cover image pixel with secret data bit imperceptibly. In addition to being less suspicious for human eyes, LSB steganography is easy to implement. There exist different versions of LSB embedding to make steganography more secure. LSB plus-minus (LSBPM) [16] also known as LSB matching, is one of the advancements in standard LSB replacement steganography. LSBPM adds or subtracts '1' from the least significant bit of cover image pixel according to secret message bit. Spatial steganography is sensitive to filtering, scaling, translation, rotation and cropping on stego image. For increasing the security, LSB method can be combined with other steganography methods. Spatial steganography is further classified into sequential and random embedding. Sequential embedding starts from the first bit of the cover image and sequentially proceeds until no secret bit is left. On the contrary, random embedding embeds the secret message bits randomly in the cover image, hence it increases security by providing no fixed pattern for detection.

Frequency transformation based steganography overcomes the limitations of spatial based techniques. In frequency based techniques, the cover image is transformed from spatial to frequency domain and the coefficients of transform are modified to embed secret data. Hence, transform domain based steganography is less perceptible with less capacity, though it complicates the implementation. Some examples of frequency based steganography are DCT (Discrete Cosine Transformation), DWT (Discrete Wavelet Transformation), and DFT (Discrete Fourier Transformation) based steganography. In DCT based techniques, the image is divided into 8 × 8 blocks followed by quantization of each block. Later, quantized components are used for embedding, finally inverse DCT is applied to get back the image. Different types of steganography and steganalysis techniques are illustrated in Fig. 1. In this paper, LSB steganography is discussed for its security with attacks.

Steganography affects perceptional and statistical properties of the image. Embedding can be noticed by comparing pixel values and the file size of the original and stego images. Statistical properties consist of the mean value, standard deviation, histogram modification, etc. Due to the addition of information, steganography affects statistical properties internally. So, steganalysis uses the embedding side-effects to detect their existence.

Steganalysis is orthogonal to steganography which targets affected image properties to enhance steganography security. Steganalysis reveals the secret communication by inspecting suspected image [2, 6, 9, 10]. Advancement in steganography also results in a new steganalysis technique, hence both the fields are explored

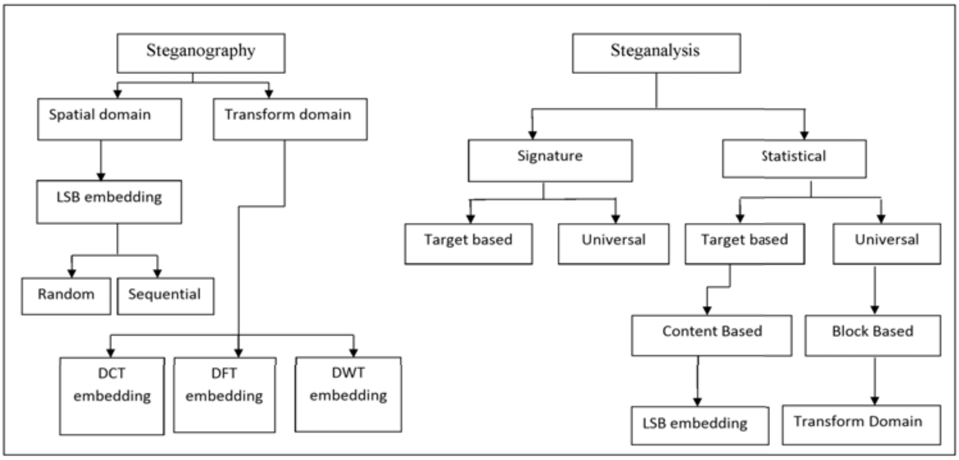


Figure 1: Steganography and Steganalysis classification [18]

by researchers equally. Based on information about steganography technique and the original image, steganalysis is categorized into two classes: blind steganalysis and targeted steganalysis. In blind steganalysis, no information of the cover image and steganography technique is available which makes it comparatively more challenging and realistic than that of targeted steganalysis [3, 13, 15].

In targeted steganalysis, embedding method with original image is provided along with given stego image. It depicts better detection with no random assumptions [8, 9, 17]. On the basis of applied steganography technique, targeted steganalysis is further divided into two categories: content based steganalysis and block based steganalysis [18]. Content based steganalysis produces a detector for LSB based steganography [22]. Block based steganalysis develops the method to detect transform based steganography. Here, blocks can vary in size and each block can adopt a different method for secrecy detection [4, 7, 14, 21].

In content based steganalysis, spatial domain based steganography techniques are targeted. Since LSB steganography modifies the direct pixel values, detection method inspects the modified pixels and related properties. Fridrich [19] used first and second order Markov chains for imitating the difference between adjacent pixels before and after LSB embedding. Resultant difference matrix acted as a feature set for classification using SVM. In this paper, we work on content based steganalysis for imperceptibility and embedding capacity.

Steganography methods can be considered as statistically affected and non-statistically affected methods. Non-statistically affected steganography techniques embed data without affecting the statistical properties of the image, while statistically affected steganography modifies the statistical properties of the original image. In this paper, we are targeting statistically affected LSB embedding methods by considering information theoretic features as steganalysis measures. First, the stego image is processed through the feature extraction phase. Next, entropy and joint entropy features are selected for discriminating original image from the stego image. The proposed method is also analyzed for reliability by attacking the stego image. We use a dataset of grayscale images for all of our experiments. Experimental results show comparison of entropy and joint entropy for steganalysis. For classification purpose, we use entropy and joint entropy features to train multiple classifiers. Based on detection accuracy, we choose SVM and ensemble classifiers for final results.

Rest of the paper is organized as follows. A survey of the related work is included in Section 2. Section 3 gives an introduction of LSB embedding and information theoretic metrics. Section 4 explains the information theoretic measures for image steganalysis and their effects as used in this work. Section 5 presents the performance evaluation and experimental results. Finally, the work is concluded in Section 6.

2 Literature Survey

Existing research in targeted steganalysis has extracted features from the suspected and original images individually. In recent years, the number of features has increased for advanced and accurate detection. Pevny et al. [19] proposed first and second order Markov chains for imitating the difference between adjacent pixels after and before LSB embedding. Resultant difference matrix acted as a feature set to classify cover image from stego image using SVM.

Fillatre & Lionel [5] added on by introducing an asymptotically and uniformly more powerful test to find out the hidden message bits irrespective of the hiding ratio. This technique explored the parametric model of the natural images where physical dependency between pixels is exploited. Generally, it sets the upper bound of hidden message bits for LSB embedding. Lerch-Hostalot & Megías [12] proposed the LSB detection idea by comparing the correlation of modified and adjacent pixels prior and post-embedding. The pattern generated after analysis of stego file is compared with the pattern of pixels in the original media. Results depicted that there exists a strong dependency in pixels of natural image that gets affected after embedding.

Kodovsky & Fridrich [11] worked on detection accuracy of steganalysis scheme by preprocessing the cover image. In her proposal, downsampling of the cover image is done before applying LSB steganography technique. Observations showed that processing of cover image affects the detectability. It draws attention toward the processing of cover image before analysis by comparing directly cover and stego images.

Sadat et al. [20] introduced entropy for motion vector steganalysis. He deployed block entropy to determine the texture and precision of the motion vector to differentiate between high and low textured blocks. High textured blocks were used as effective features and used in re-estimation of the motion equation. It shows that block-entropy classifies the video into cover and stego files with more precision.

In this paper, the developed steganalysis methods exploit features of stego and cover images for comparison. An enhanced technique can consider joint information from stego and cover images. In the proposed work, both original and suspected images are processed to extract a common measure for detection by using information theory. Here, LSB steganography is targeted with entropy and joint entropy for extracting information. Reliability of both metrics is analyzed by further attacking the image. It is shown that joint entropy is an appropriate steganalysis measure in comparison to entropy. Further extracted information is fed into classifiers. From experiments, we infer that Support Vector Machine (SVM) and ensemble classifiers give better detection accuracy. Finally, We conclude that joint entropy improves detection by using information from both the original and the stego images.

3 Background

3.1 LSB Embedding

LSB embedding is the most straightforward and standard steganography method that influences the content of an image by modifying the least significant bits of each pixel. Consider an 8-bit grayscale image I_c with $M \times N$ pixels and a secret message S with n bits to be communicated secretly by I_c, over a local channel in an imperceptible way. Let the last m bits of the cover image be replaced according to the message length and the quality requirements of the stego image, where

$$I_c = \{x_{ij} \mid \quad 0 \leq i < M, \quad 0 \leq j < N\}, \tag{1}$$

$$x_{ij} \;\in\; \{0, 1, \ldots, 255\}, \quad \textit{and}$$

$$S = \{s_k \mid \quad 0 \leq k < n, \; s_k \in \{0, 1\}\}. \tag{2}$$

First, S should be compatible to a cover image so that the quality of the image can be maintained after embedding. Therefore, the rightmost bit of the pixels should be used for embedding. However, the quality and embedding capacity requirements during steganography need more embedding bits with less perceptibility. Thus for embedding, we consider m rightmost bits from each selected pixels, and the value of m depends on the requirement of the method. Generally, the preferred value of m is less than 4, as up to there, quality gets affected in imperceptible range. For this purpose, S is rearranged to form a virtual image I' compatible to the original image, such that,

$$I' = \{y_i \mid \quad 0 \leq i < n'\}, \tag{3}$$

where,

$$y_i \in \{0, 1, \ldots, 2^m - 1\},$$

$n' < M \times N$, and I' contains non binary values for $m > 1$. Binary message S is mapped to non binary message I' by following equation:

$$y_i = \sum_{k=0}^{m-1} S_{i \times m+k} \ \cdot \ 2^{m-1-k}. \tag{4}$$

Now, embedding of I' is done by selecting x_l from I_c and replacing m least significant bits with secret message bits to form z_i as:

$$z_i = x_l \ - \ x_l \ mod \ 2^m, \tag{5}$$

where z_i is a pixel of the stego image that keeps secret bits without revealing their existence and I_s is the corresponding stego image. At the receiving end, the same process is reversed to get the secret message bits, ss_i back with the image as:

$$ss_i = z_l \ mod \ 2^m. \tag{6}$$

In stego image, z_l pixels are selected that are modified during steganography. From the selected pixels of the stego image, m LSB bits are extracted and arranged in an informative form similar to a secret message.

3.2 Information Theory

Information entropy defines the uncertainty and predictability of a discrete system. Let X be a discrete random variable with D as its domain with probability mass function as:

$$P(X) = \sum_{x \in D} p(x), \tag{7}$$

then, entropy of the system X can be defined mathematically as:

$$H(X) = -\sum_{x \in D} p(x) \ log \ p(x). \tag{8}$$

Here, logarithm of base 2 is taken as information is revealed in binary form. Entropy represents information in bits. Mathematical representation of entropy shows that it is inversely proportional to probability mass function of the events in the system. For an instance, if an event has maximum probability then it will reveal less information with more certainty. Hence, entropy depends on probability of the event. In terms of expectation $E(X)$ of the system, entropy is defined as:

$$H(X) = -E(X) \ log \ P(X). \tag{9}$$

Uncertainty associated with a set of random variables is termed as joint entropy. Joint entropy reveals information by considering more than one random variable at the same time. Let X and Y are two discrete random variables with U and V

as their respective domains having joint distribution $P(X,Y)$. Then, joint entropy $H(X,Y)$ for pair (X,Y) is defined as:

$$H(X,Y) = -\sum_{x\in U}\sum_{y\in V} p(x,y)\ log\ p(x,y). \tag{10}$$

Also, when we have expectation value by observing X and Y together, then joint entropy can be defined as:

$$H(X,Y) = -E(X,Y)\ log\ P(X,Y). \tag{11}$$

The joint entropy can also be formulated in terms of individual entropies of X and Y, i.e., $H(X)$ and $H(Y)$ respectively, with conditional entropy. The conditional entropy, $H(X|Y)$ of X and Y is the amount of information about X with prior information of Y. Then, joint entropy is defined as:

$$H(X,Y) = H(Y) + H(X|Y) = H(X) + H(Y|X). \tag{12}$$

In case of independent random variables, $H(X|Y) = H(Y|X) = 0$, as there is no benefit of prior information from another variable. Therefore,

$$H(X,Y) \leq H(X) + H(Y) \geq 0, \tag{13}$$

and,

$$H(X,Y) \geq max(H(X), H(Y)) \tag{14}$$

In this paper, random variables are related to an image system in order to compute entropy and joint entropy of the system. An image is a bundle of pixels in a fixed pattern to deliver related information. Image entropy extracts information of its uniformity and randomness. For an individual image, entropy reveals the information needed while joint entropy is used for knowing the information about the images at the same time. Here, entropy and joint entropy of the images are computed to get related information for detecting LSB steganography. Joint entropy needs joint probability distribution for its computation that requires joint histogram of the participating images. The joint histogram is a 2-dimensional representation of 1-dimensional histograms, where, the first dimension is for the intensity of one image and second dimension for another image. So, instead of evaluating detection measures separately for the original and stego images, joint entropy extracts combined information from both the images. Hence, we speculate that entropy and joint entropy can be used to discriminate the stego image from the cover image.

4 Proposed Information Theoretic Steganalysis

In this section, steganalysis for LSB steganography is proposed with information theoretic metrics. The proposed system works in two phases: first is embedding & training phase, followed by the second phase of verification. These two phases are described in subsequent subsections. Block diagram of the proposed information theoretic image steganalysis is shown in Fig 2.

- ❖ Embedding & Training Phase: For communicating secretly, a secret file is needed, which is to be embedded in a media file. Initially, an image and a text file as a secret message are given as input for LSB steganography system. After embedding, steganalysis scheme extracts entropy and joint entropy of the embedded image. Next, according to the behavior of the extracted features, we train the classifier for discriminating stego image from the cover image. We consider the SVM and ensemble classifiers for the proposed scheme. We conduct experiments with suspected image for entropy and joint entropy-based measures for detection of the existence of the steganography.

  Empirical results show the increase in original joint entropy during the training phase that indicates the existence of steganography. On the other hand entropy showed no appreciable change in its original value. Thus, joint entropy is an effective measure for steganalysis in comparison to entropy.

- ❖ Verification Phase. In this phase, system inspects the suspected image by scheme A and scheme B. In scheme A, entropy and joint entropy are used to detect steganography by trained classifiers. Results show that joint entropy can easily detect embedding while entropy shows no significant participation in steganalysis. In scheme B, the histogram of the given image is attacked. After that resultant image is analyzed by training classifiers for entropy and joint entropy measures thus making a final decision whether steganography is still detectable or not. According to the experimental results of scheme B, joint entropy is proved to be an effective measure for detection also with the attack on stego image, while entropy shows no change in its behavior.

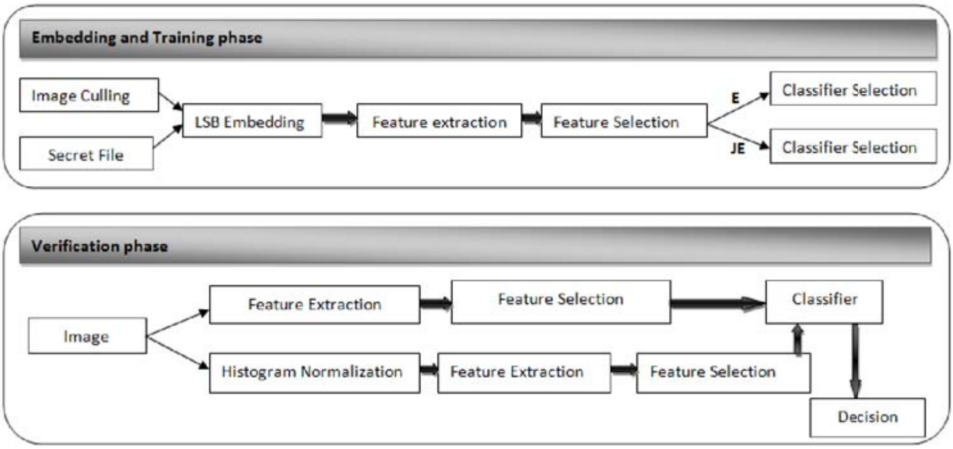


Figure 2: The proposed system framework

It is worthwhile to compare information theoretic measures for image steganalysis with individual feature based steganalysis. Traditional steganalysis focuses on image size, statistical properties, energy and contrast of the image for detection of steganography, while we introduce entropy and joint entropy for detection purpose. For analyzing the robustness of proposed metrics, histogram of the stego image is modified and then entropy and joint entropy features of the image are used for steganalysis. Further, classifiers are trained and tested with the proposed measures for efficient detection.

4.1 Embedding & Training Phase

Adaptation of steganography technique ensures the embedding of secret message in the cover image. Here, the cover image file is embedded with secret file by using LSB steganography method. In training phase, the proposed steganalysis targets applied steganography for its detection. First, features of the suspected (stego) image are extracted followed by relevant feature selection for detection. Next, we train the classifiers according to the decided discrimination function. For the proposed method, entropy and joint entropy metrics are selected during feature selection.

- *Entropy*: Entropy is one of the features that may get affected by embedding. Since information is added into the cover image by modifying its original contents, so randomness may either increase or decrease. However, steganography produces stego image perceptibly similar to cover image, and entropy is the factor that may affect during the embedding phase. Thus, we apply entropy metric over steganography to detect its existence.

- *Joint Entropy*: While entropy of the image can be extracted for steganalysis, joint entropy of stego image with the cover image can be a reliable metric. Since joint entropy considers stego and cover images together for the joint information, this may distinguish stego and cover images more clearly. After analysis with entropy and joint entropy, experimental results show that joint entropy gives a fixed pattern for different hiding ratio. Hence, joint entropy performed better than entropy, as joint entropy gives results with a fewer false alarms. Joint entropy does not change its behavior after the attack.

Above, feature selection process is followed by a classifier selection and training stage. We experimented with Fischer linear discriminate (FLD), logistic, support vector machine (SVM) and ensemble classifiers, the last two gave better detection accuracy.

4.2 Verification Phase

For a given steganography method, a secret message is embedded into a cover image to get cover/stego set. In the proposed steganalysis technique, a cover/stego pair is processed through two different schemes as described below. Here, the verification phase includes testing of the given image with and without attack. The verification phase aims to check the given image pair with entropy and joint entropy metrics of the proposed schemes.

- *Scheme A*: Verification with entropy and joint entropy. One perceptive way is to feed the image pair into a trained classifier using the selected features. First, choose a cover/stego pair and extract entropy and joint entropy as features. Next, based on information entropy (IE), a discrimination function D_e is formulated for the classifier to discriminate between cover image I_c and stego image I_s as:

$$D_e = \|IE(I_c) - IE(I_s)\|. \tag{15}$$

Instead of evaluating individual entropy of cover and stego images, joint entropy of both the images is calculated. Joint entropy (JE) of cover and stego images is proposed here for another discrimination function D_j for classifier as:

$$D_j = \|JE(I_c, I_c) - JE(I_c, I_s)\|. \tag{16}$$

The given image pair is checked with both the discrimination functions, entropy and joint entropy. It is observed that the joint entropy metric gives better detection.

- *Scheme B*: Here, we verify reliability of entropy and joint entropy by attacking the image histogram. The stego image is modified by histogram normalization, and then reliability is inspected by a respective discrimination function. Histogram normalization works as:

$$I_{s\_norm} = N(I_s) = \sum_{j=0}^{k} pr(n_j), \quad k = 0 \ldots L, \tag{17}$$

where,

$$n_j = \frac{j}{L},$$

and k is the maximum intensity level used in the image. L is the upper bound of intensity level. I_{s_norm} is a normalized image, n_j is a normalized intensity level of the input image, $pr(n_j)$ is probability density function of the image with normalized levels. For example, as shown in Fig. 3, stego image is normalized by the given equations that result in Fig. 3(c-d). As in scheme A, the modified image is analyzed by entropy and joint entropy metrics, respectively, as given in the following two equations:

$$D_{he} = \|IE(I_c) - IE(I_{s\_norm})\|, \tag{18}$$

and,

$$D_{hj} = \|JE(I_c, I_c) - JE(I_c, I_{s\_norm})\|. \tag{19}$$

Verification phase provides discrimination functions for classifiers to decide whether the given image is stego or original. According to the proposed discrimination functions after histogram attack, D_{he} for entropy metric and D_{hj} for joint entropy metric results show that joint entropy is still able to distinguish cover and stego images.

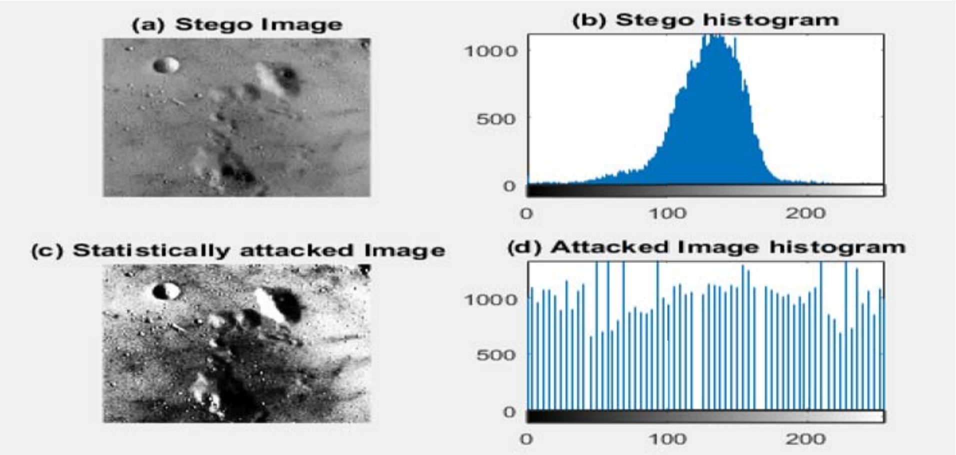


Figure 3: Example of a sample stego image and its normalized image with histogram: (a) Original stego image, (b) Histogram of the original stego image, (c) Stego image after histogram normalization, i.e., normalized stego image, a.k.a. the stego image after histogram attack, and (d) Histogram of the normalized stego image.

5 Performance Evaluation

The performance of information theoretic steganalysis is studied in this section. We compare both the proposed metrics – entropy and joint entropy – and then select joint entropy as a better metric for steganalysis. Therefore, we include initially results for both the entropy and joint entropy, and later we include detailed results for joint entropy only. We carry out experimental results using different secret files and binary classifiers for detection.

The performance of the steganalysis method is measured by detection accuracy rate with respect to the specific steganography:

$$P_{detect} = 1 - D_{error}, \tag{20}$$

where D_{error} is the probability of error difference. When we target a steganography method, the error difference between actual embedding and detected embedding acts as a steganalysis measure. Error difference is defined as a difference between original embedding features, o_{embed} and detected embedding features, d_{embed} :

$$D_{error} = o_{embed} - d_{embed}. \tag{21}$$

An ideal steganalysis technique has zero error difference with 100% accuracy. Detected embedding decision process includes false positives and false negatives. Targeted steganalysis tries to maximize detection accuracy by observing these two aspects. False positives give the false alarm by placing the cover image into stego image class, while false negatives contribute to false alarm by escaping stego image undetectable. So, extracted embedding during steganalysis scheme d_{embed} can be defined as:

$$d_{embed} = \frac{1}{2}(fp + fn), \tag{22}$$

where fp and fn are respective false positives and false negatives of steganalysis scheme. Therefore detection accuracy rate is:

$$P_{detect} = 1 - [o_{embed} - (\frac{1}{2}(fp + fn))]. \tag{23}$$

5.1 Experimental Setup

For experiments, we have used MATLAB 2017a Windows 7, CPU core i7 with 10 GB of RAM. We have examined 1000 grayscale images of dimensions 256×256, 384×512, 512×512 and 1024×1024 for training and testing. The dataset includes benchmark images taken from nature and some random captures, which are used in image processing and data hiding. A few images used from this image dataset, are shown in Fig. 4. For analyzing the effects of secret file size, we have taken 1000, 5000, 10000 and 20000 KB text files. Each image of the dataset is processed by LSB embedding methods with various secret text files.

After obtaining the stego and cover images, all the possible features are extracted, including entropy and joint entropy. Further, with extracted information theoretic features, every stego image is inspected for maximum embedding detection. We have used binary classifiers for decision making between the cover and stego images. For the initial training phase, we have used half of the images of the dataset and rest half for testing process. For further experiments, we have used a different ratio of training and testing sets from the dataset to examine the accuracy of the corresponding classifier. During the training and testing phase, the extracted features are compared with the original image while the verification phase checks the reliability of the proposed metrics by attacking the histogram of the image.

5.2 Performance Analysis of Proposed Metrics

For performance analysis of the proposed steganalysis features, we gather 1000 gray-scale images of different dimensions and use them as cover images. The spatial domain based steganography methods, namely, standard LSB embedding and LSB plus-minus (LSBPM) [16] are used to create corresponding stego image datasets. Furthermore, the produced stego images are inspected with entropy and joint entropy based features.

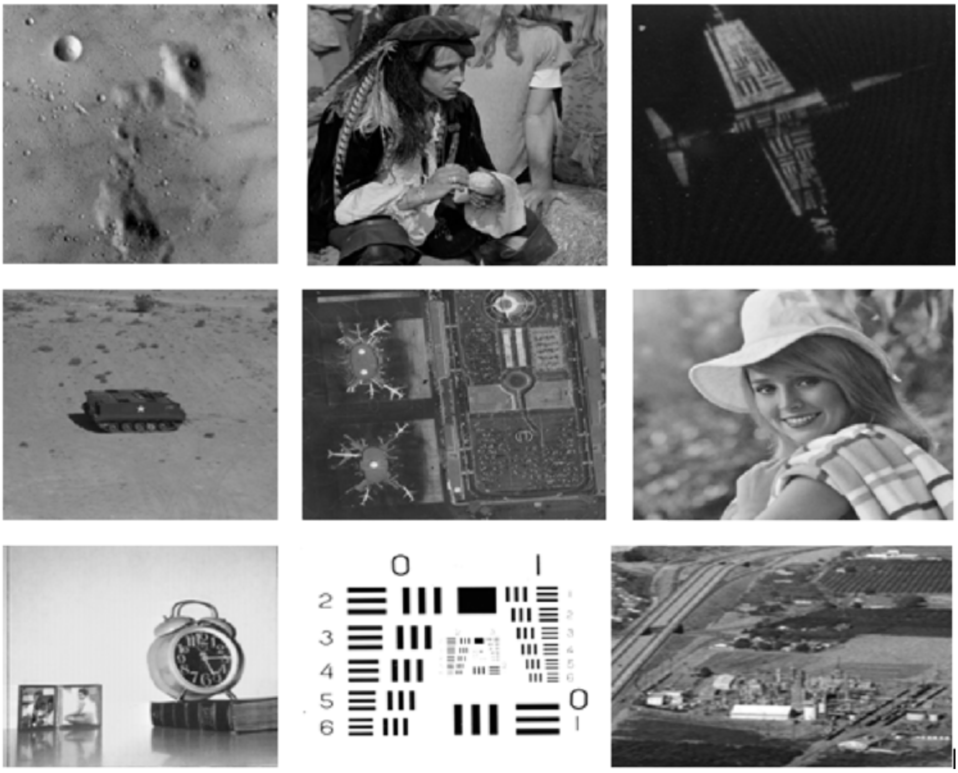


Figure 4: A few gray scale images taken from the dataset.

5.2.1 Entropy and joint entropy with varying hiding ratio

Both entropy and joint entropy are information theoretic metrics, however, the information gain during steganalysis makes the difference. From equations (15), (16), (18) and (19), we compare the performance of these metrics as:

$$P_{compare} = D_j - D_e, \tag{24}$$

or

$$P_{compare} = D_{hj} - D_{he}. \tag{25}$$

Here, the accuracy and efficiency of the proposed metrics are analyzed for detecting LSB and LSBPM steganography methods. First, the stego image generated by steganography methods is experimented for detection purpose by using entropy and joint entropy metrics. We repeated experiments over a set of twenty images selected randomly. Since the behavior remains the same throughout our experiments, hence we represent the results of one representative sample of our experiments in Table 1 due to space consideration. Table 1 represents the results obtained for entropy and joint entropy based features for detecting standard LSB embedding and LSBPM steganography methods.

Table 1: Entropy and joint entropy for LSB and LSBPM steganography methods over a randomly selected image.

steganograpy	Bit Rate	Entropy	Joint entropy
LSB	0.20	6.709312	6.709312
	0.45	6.709350	6.749539
	0.73	6.709482	7.700813
	0.99	6.709482	7.700813
LSBPM	0.20	6.756524	6.701335
	0.45	6.761320	6.790011
	0.73	6.774975	7.713119
	0.99	6.786145	7.895221

From Table 1, we observe that the standard LSB embedding method does not show any significant variation in entropy values, while joint entropy increases with increase in bit rate. Almost similar is the case with LSBPM steganography, in which the entropy values are increased very marginally, while joint entropy increases significantly over increase in bit-rate. We also got the similar pattern of variations of entropy and joint entropy with varying bit-rates after histogram attack. This is discussed in the next sub-section. Therefore, we conclude from these results that joint entropy is an effective measure over entropy for detection.

5.2.2 Effect of histogram attack

Further, performance of entropy and joint entropy is explored by attacking histograms of the stego images. Here, the stego images of used dataset with histogram attack are experimented to check the metrics efficiency. Table 2 shows a representative sample result for showing the reliability of proposed features for detecting LSB steganography methods after histogram attack.

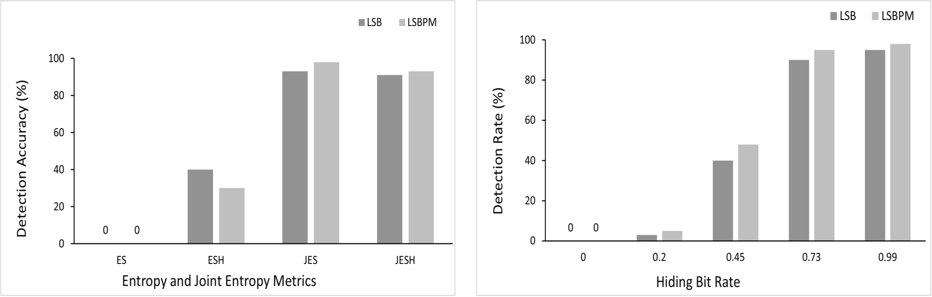
Table 2: Entropy and joint entropy for LSB and LSBPM steganography methods after histogram attack over a randomly selected image.

steganograpy	Bit Rate	Entropy	Joint entropy
LSB	0.20	5.921757	6.709312
	0.45	5.921710	6.743879
	0.73	5.918913	7.475977
	0.99	5.918913	7.475977
LSBPM	0.20	5.935647	6.712546
	0.45	5.938745	6.735481
	0.73	5.949965	7.328165
	0.99	5.949965	7.452096

From Table 2, we observe the similar behaviour as in Table 1, that there is no appreciable change in entropy values with varying bit-rates, while joint entropy increases with increase in hiding bit-rates. From these experimental results, taken over a range of grey-level images with varying hiding rates and histogram attacks, we conclude that (i) entropy is not a discriminative feature, while (ii) joint entropy can be used reliably for both the LSB and LSBPM steganography methods with varying hiding ratio and with histogram attacks.

Next, we experiment both the metrics for detection accuracy before and after the statistical, i.e., histogram attack. We calculate detection accuracy with entropy and joint entropy, before and after histogram attack, for both LSB and LSBPM methods. The entropy metric before and after histogram attack does not indicate any pattern for detecting steganography. Thus, the detection accuracy with entropy for both the methods, i.e., before and after attacks, is quite low.

However, the joint entropy metric deflects after embedding in both cases, i.e., before and after attack. The stego image has marginally higher joint entropy with respect to the original image. The detection accuracy with joint entropy is found to be above 90% for all the cases across all the randomly selected images. The results are depicted in Fig. 5(a) for both entropy and joint entropy before and after histogram attack.



(a) (b)

Figure 5: (a) Detection accuracy with entropy and joint entropy (before and after attack): ES (Entropy of stego image), ESH (Entropy of stego image after histogram attack), JES (Joint entropy of stego image), JESH (Joint entropy of stego image after histogram attack), and (b) Detection rate by joint entropy with varying hiding rate for LSB and LSBPM steganography methods.

All the results presented herein above in Tables 1 & 2 and Fig. 5(a), discard use of entropy as an information theoretic feature for steganalysis. Next, we calculate joint entropy with different hiding ratios for both LSB and LSBPM methods, and plot in Fig. 5(b). These plots show that joint entropy varies significantly with different hiding ratios. Thus, joint entropy is shown to be an effective metric for detection of steganography with histogram attack. All the above results computed for joint entropy, finally, establish joint entropy as a distinctive feature for steganalysis.

Table 3: Joint entropy with different classifiers.

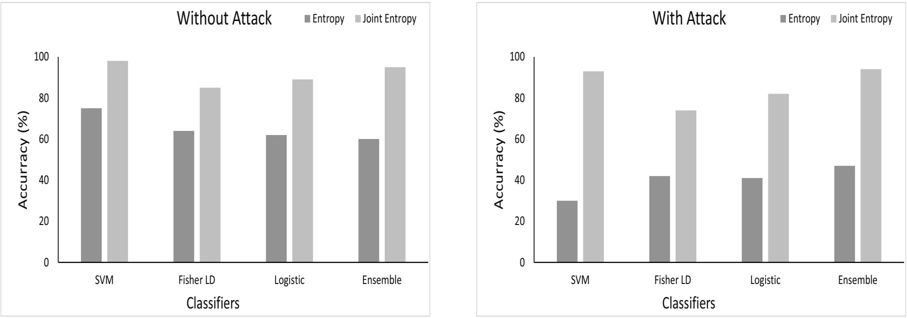
Classifier	Detection			
	Correct	Incorrect	Total	Accuracy
SVM	245	5	250	0.98
Fisher LD	454	78	535	0.85
Logistic	135	17	152	0.89
Ensemble	570	32	602	0.95

5.2.3 Detection accuracy with classifiers

We use two class classifiers for discrimination between cover and stego images. For classification, we selected four classifiers, namely, Support Vector Machine (SVM), Fisher linear discriminant (Fisher LD), logistic, and ensemble classifiers. Table 3 includes the results of joint entropy with these four classifiers before the attack. The SVM and ensemble classifiers have shown higher accuracy over the Fisher LD and logistic classifiers. The corresponding classification results of joint entropy, after histogram attack, are shown in Table 4. In analogy with the classification results before attack, the results of SVM and ensemble classifiers, after attack, have higher accuracy over the other two classifiers.

Table 4: Joint entropy after histogram attack with different classifiers.

Classifier	Detection			
	Correct	Incorrect	Total	Accuracy
SVM	232	18	250	0.93
Fisher LD	396	139	535	0.74
Logistic	124	28	152	0.82
Ensemble	566	36	602	0.94

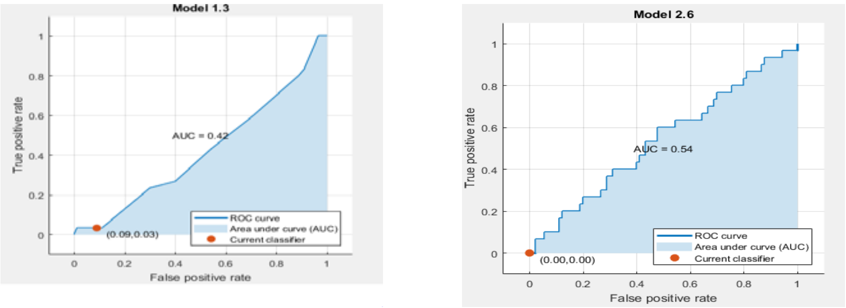


(a) (b)

Figure 6: Classification results using entropy and joint entropy: (a) before attack, and (b) after attack.

The corresponding classification results of Tables 3 & 4 are plotted in Fig. 6; the plots in Fig. 6(a) represent classification accuracy before attack, and plots in Fig. 6(b) represent the results after attack. In addition, the plots in Fig. 6 also include classification accuracy with entropy. It can be observed that the detection accuracy is much higher with joint entropy metric over entropy. Moreover, the detection accuracy with entropy decreases very significantly after the statistical histogram attack. For all classifiers, the classification accuracy with joint entropy, before and after attack, remains above 80%, this is above 90% with the state of art classifiers, namely, SVM and ensemble classifiers.

Finally, we plot ROC curves for joint entropy with respect to the steganography detection before and after the attack in Fig. 7. These experiments are done by averaging observations from LSB and LSBPM steganography.



(a) (b)

Figure 7: ROC curve with joint entropy: (a) without attack, and (b) with attack.

In summary, the steganalysis performance with information theoretic metrics, namely, entropy and joint entropy, for detecting LSB class of steganography, is shown by Table 1 to Table 4 and Fig. 5 to Fig. 7. As the embedding rate increases, joint entropy is shown to be an effective detection metric with a better accuracy rate in comparison to entropy. The detection rate with entropy metric, before the attack, is 70%, and after histogram attack, it decreases to 30%. In contrast, the same detection rate with joint entropy metric before the attack is 98%, and remains almost the same after the attack, it reduces very marginally to 93%. Therefore, the detection of LSB steganography methods by joint entropy metric is much superior than that of entropy metric.

6 Conclusion

In this work, information theoretic measures for steganalysis are proposed for detecting LSB steganography methods. Existing steganalysis schemes distinguish the cover and stego images based on their individual information. In the proposed scheme, images with their original versions are used to extract entropy and joint entropy based features. Extracted features were used to distinguish stego and cover images by using SVM and ensemble classifiers. For checking the efficiency of the method, images were attacked statistically and detection accuracy was then measured. Experimental results showed that entropy is not a reliable measure for the detection of LSB steganography, while joint entropy is shown to be quite discriminative.

In order to analyze the proposed detection measures for different hiding ratio, we have examined entropy and joint entropy over different secret files. Detection accuracy by joint entropy is observed to be better than the detection accuracy by entropy. For classification between cover and stego images, a few classifiers are compared in terms of false alarms, correct prediction and accuracy; the SVM and ensemble classifiers show maximum accuracy.

Future research augmentation can be to use the proposed metrics for detecting frequency based steganography. Further analysis can be done with other attacks and observing their effects on detection accuracy. In order to define a standard steganalysis measure, one may work to derive a well-formed metric using joint entropy and compare its performance with other well-known steganalysis methods.

Acknowledgements

The authors thank Managing Editor Boglárka G.-Tóth, Associate Editor Zoltan Kato, and anonymous reviewers for their help and bringing clarity in presentation of this work. The author, Rajeev Kumar, also acknowledges help of Roopam Sadh, Neeraj Pathak, and Priti Kumari for carrying-out some of the experimental results for major revision of this manuscript.

References

[1] Amin, Muhalim Mohamed, Salleh, Mazleena, Ibrahim, Subariah, Katmin, Mohd Rozi, and Shamsuddin, MZI. Information hiding using steganography. In *Proc. 4th Nat. Conf. Telecommunication Technology (NCTT)*, pages 21–25. IEEE, 2003. DOI: `10.1109/NCTT.2003.1188294`.

[2] Chandramouli, Rajarathnam, Kharrazi, Mehdi, and Memon, Nasir. Image steganography and steganalysis: Concepts and practice. In *Proc. Int. Workshop Digital Watermarking (IWDW), LNCS, vol. 2939*, pages 35–49. Springer, 2003. DOI: `10.1007/978-3-540-24624-4_3`.

[3] Chen, Xiaochuan, Wang, Yunhong, Tan, Tieniu, and Guo, Lei. Blind image steganalysis based on statistical analysis of empirical matrix. In *Proc. 18th Int. Conf. Pattern Recognition (ICPR)*, volume 3, pages 1107–1110. IEEE, 2006. DOI: `10.1109/ICPR.2006.332`.

[4] Do, Minh N and Vetterli, Martin. Wavelet-based texture retrieval using generalized gaussian density and kullback-leibler distance. *IEEE Trans. Image Processing*, 11(2):146–158, 2002. DOI: 10.1109/83.982822.

[5] Fillatre, Lionel. Adaptive steganalysis of least significant bit replacement in grayscale natural images. *IEEE Trans. Signal Processing*, 60(2):556–569, 2011. DOI: 10.1109/TSP.2011.2174231.

[6] Fridrich, Jessica, Goljan, Miroslav, and Du, Rui. Detecting LSB steganography in color, and gray-scale images. *IEEE Multimedia*, 8(4):22–28, 2001. DOI: 10.1109/93.959097.

[7] Fu, Dongdong, Shi, Yun Q, Zou, Dekun, and Xuan, Guorong. Jpeg steganalysis using empirical transition matrix in block DCT domain. In *Proc. IEEE Workshop Multimedia Signal Processing*, pages 310–313. IEEE, 2006. DOI: 10.1109/MMSP.2006.285320.

[8] Hawi, Tariq Al, Qutayri, MA, and Barada, Hassan. Steganalysis attacks on stego-images using stego-signatures and statistical image properties. In *Proc. IEEE Region 10 Conf. TENCON*, pages 104–107. IEEE, 2004. DOI: 10.1109/TENCON.2004.1414542.

[9] Johnson, Neil F and Jajodia, Sushil. Steganalysis of images created using current steganography software. In *Proc. Int. Workshop Information Hiding (IWIH), LNCS, vol. 1525*, pages 273–289. Springer, 1998. DOI: 10.1007/3-540-49380-8_19.

[10] Johnson, Neil F and Jajodia, Sushil. Steganalysis: The investigation of hidden information. In *Proc. IEEE Information Technology Conf. Information Environment for the Future (Cat. No. 98EX228)*, pages 113–116. IEEE, 1998. DOI: 10.1109/IT.1998.713394.

[11] Kodovsky, Jan and Fridrich, Jessica. Effect of image downsampling on steganographic security. *IEEE Trans. Information Forensics & Security*, 9(5):752–762, 2014. DOI: 10.1109/TIFS.2014.2309054.

[12] Lerch-Hostalot, Daniel and Megías, David. LSB matching steganalysis based on patterns of pixel differences and random embedding. *Computers & Security*, 32:192–206, 2013. DOI: 10.1016/j.cose.2012.11.005.

[13] Lie, Wen-Nung and Lin, Guo-Shiang. A feature-based classification technique for blind image steganalysis. *IEEE Trans. Multimedia*, 7(6):1007–1020, 2005. DOI: 10.1109/TMM.2005.858377.

[14] Liu, Shaohui, Yao, Hongxun, and Gao, Wen. Steganalysis of data hiding techniques in wavelet domain. In *Proc. Int. Conf. Information Technology: Coding and Computing (ITCC)*, volume 1, pages 751–754. IEEE, 2004. DOI: 10.1109/ITCC.2004.1286558.

[15] McBride, Brent, Peterson, Gilbert, and Gustafson, Steven. A new blind method for detecting novel steganography. *Digital Investigation*, 2:50–70, 02 2005. DOI: `10.1016/j.diin.2005.01.003`.

[16] Mielikainen, Jarno. LSB matching revisited. *IEEE Signal Processing Letters*, 13(5):285–287, 2006. DOI: `10.1109/LSP.2006.870357`.

[17] Niimi, Michiharu, Eason, Richard O, Noda, Hideki, and Kawaguchi, Eiji. Intensity histogram steganalysis in BPCS-steganography. In *IS&T/SPIE Electronic Imaging*. SPIE, 2001. DOI: `10.1117/12.435440`.

[18] Nissar, Arooj and Mir, Ajaz Hussain. Classification of steganalysis techniques: A study. *Digital Signal Processing*, 20(6):1758–1770, 2010. DOI: `10.1016/j.dsp.2010.02.003`.

[19] Pevny, Tomáš, Bas, Patrick, and Fridrich, Jessica. Steganalysis by subtractive pixel adjacency matrix. *IEEE Trans. Information Forensics & Security*, 5(2):215–224, 2010. DOI: `10.1109/TIFS.2010.2045842`.

[20] Sadat, Elaheh, Faez, Karim, and Saffari Pour, Mohsen. Entropy-based video steganalysis of motion vectors. *Entropy*, 20(4):244, 2018. DOI: `10.3390/e20040244`.

[21] Sullivan, Kenneth, Bi, Zhiqiang, Madhow, Upamanyu, Chandrasekaran, Shivkumar, and Manjunath, BS. Steganalysis of quantization index modulation data hiding. In *Proc. Int. Conf. Image Processing (ICIP).*, volume 2, pages 1165–1168. IEEE, 2004. DOI: `10.1109/ICIP.2004.1419511`.

[22] Trivedi, Shalin and Chandramouli, Rajarathnam. Active steganalysis of sequential steganography. In *Proc. Security and Watermarking of Multimedia Contents V*, volume 5020, pages 123–130. Int. Society for Optics & Photonics, 2003. DOI: `10.1117/12.473115`.

Received 4th January 2019