A survey on image steganography and steganalysis

Article *in* Journal of Information Hiding and Multimedia Signal Processing · May 2011

CITATIONS
485

READS
2,842

4 authors:

Bin Li
Shenzhen University
123 PUBLICATIONS **5,933** CITATIONS
SEE PROFILE

Junhui He
Hunan University
30 PUBLICATIONS **1,034** CITATIONS
SEE PROFILE

Jiwu Huang
Shenzhen University
258 PUBLICATIONS **11,160** CITATIONS
SEE PROFILE

Y.Q. Shi
New Jersey Institute of Technology
405 PUBLICATIONS **18,458** CITATIONS
SEE PROFILE

Journal of Information Hiding and Multimedia Signal Processing ©2011 ISSN 2073-4212
Ubiquitous International Volume **2**, Number **2**, April **2011**

A Survey on Image Steganography and Steganalysis

Bin Li

College of Information Engineering,Shenzhen University
No. 3688 Nan Hai Road, Shenzhen518060, China
libin@szu.edu.cn

Junhui He

School of Computer Science and Engineering
South China University of Technology
Guangzhou 510641,China
hejh@scut.edu.cn

Jiwu Huang

School of Information Science and Technology
Sun Yat-sen University
Guangzhou 510275, China
isshjw@mail.sysu.edu.cn

Yun Qing Shi

Department of Electrical and Computer Engineering
New Jersey Institute of Technology
Newark, NJ 07102, USA
shi@njit.edu

Received July 2010; revised October 2010

Abstract. *Steganography and steganalysis are important topics in information hiding. Steganography refers to the technology of hiding data into digital media without drawing any suspicion, while steganalysis is the art of detecting the presence of steganography. This paper provides a survey on steganography and steganalysis for digital images, mainly covering the fundamental concepts, the progress of steganographic methods for images in spatial representation and in JPEG format, and the development of the corresponding steganalytic schemes. Some commonly used strategies for improving steganographic security and enhancing steganalytic capability are summarized and possible research trends are discussed.*
Keywords: Digital image, information hiding, steganalysis, steganography

1. **Introduction.** Cryptography is often used to protect information secrecy through making messages illegible. However, indecipherable messages may raise an opponent's suspicion and probably lead to his destruction of such a communication manner. Therefore, steganography [1] gets a role on the stage of information security. Steganography refers to the technique of hiding information in digital media in order to conceal the existence of the information. The media with and without hidden information are called stego media and cover media, respectively [2]. Steganography can meet both legal and illegal interests. For example, civilians may use it for protecting privacy while terrorists may use it for spreading terroristic information. Compared to digital watermarking, another branch of information hiding, steganography stresses more on preserving the secrecy of

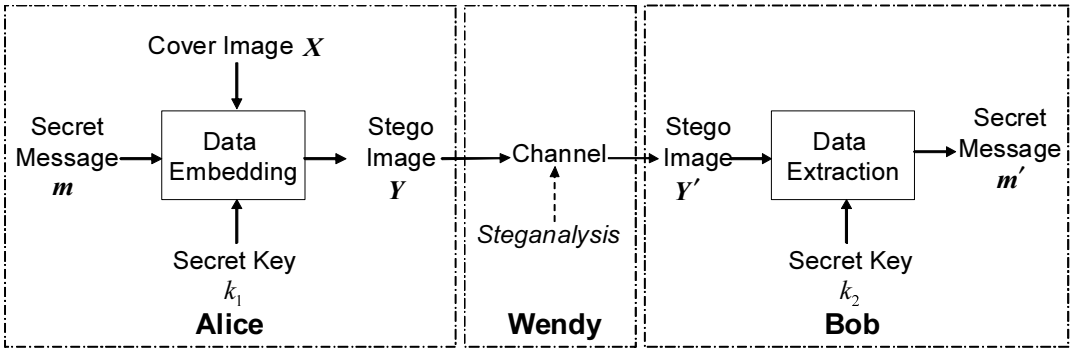


FIGURE 1. The model of steganography and steganalysis

the information instead of making the hidden information robust to attacks. For more details on the difference between steganography and digital watermarking please refer to ref. [3].

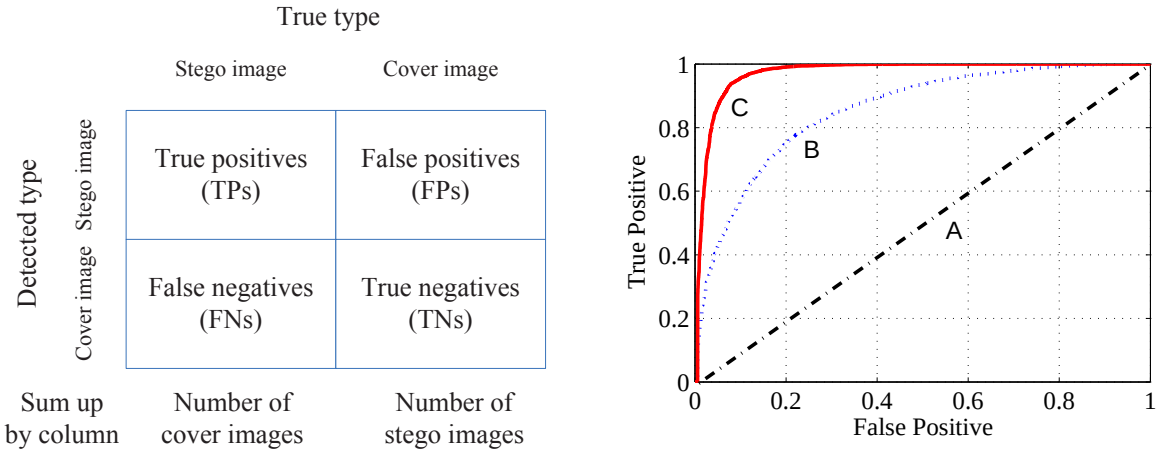
Steganalysis[4], from an opponent's perspective, is an art of deterring covert communications while avoiding affecting the innocent ones. Its basic requirement is to determine accurately whether a secret message is hidden in the testing medium. Further requirements may include judging the type of the steganography, estimating the rough length of the message, or even extracting the hidden message. Steganography and steganalysis are in a hide-and-seek game [5]. They try to defeat each other and also develop with each other.

Digital images have high degree of redundancy in representation and pervasive applications in daily life, thus appealing for hiding data. As a result, the past decade has seen growing interests in researches on image steganography and image steganalysis [3, 4, 5, 6]. This paper aims to provide a comprehensive review on different kinds of steganographic schemes and possible steganalytic methods for digital images.

The organization of this paper is as follows. In the next section, we revisit the basic model of steganography and steganalysis and their evaluation criteria. Then, in Section 3, we review some major steganography for images in spatial representation and in JPEG format. Steganalytic schemes targeted to the mentioned steganographic methods as well as some steganalytic features effective to attacking a broad class of steganography are presented in Section 4. The latest effective and commonly used techniques in steganography and steganalysis are discussed in Section 5. This paper shows some possible future research directions and concludes in Section 6.

2. **Fundamental Concepts.**

2.1. **Basic Model.** The issue in steganography and steganalysis is often modeled by the prisoner's problem [7] which involves three parties, as illustrated in Figure 1. Alice and Bob are two prisoners who collaborate to hatch an escape plan while their communications will be monitored by a warden, Wendy. Using a data embedding method $\Psi(\cdot)$, secret information $\boldsymbol{m}$ is supposed to be hidden into a cover medium $\boldsymbol{X}$ by Alice with a key k_1. The generation of an innocuous-looking stego medium $\boldsymbol{Y}$ can be described as $\boldsymbol{Y} = \Psi(\boldsymbol{X}, \boldsymbol{m}, k_1)$. On the receiver's side, the medium obtained by Bob, denoted by $\boldsymbol{Y}'$, is passed to a data extraction method $\Phi(\cdot)$ to extract information $\boldsymbol{m}'$ with a key k_2. The extraction process may be described as $\boldsymbol{m}' = \Phi(\boldsymbol{Y}', k_2)$. The steganographic scheme should ensure $\boldsymbol{m}' = \boldsymbol{m}$. Although the public key steganographic scheme is considered in some literatures, the private key steganographic scheme, where $k_1 = k_2$ is assumed, remains the most common scenario in a steganographic system. Wendy can be active or



(a) Confusion matrix (b) ROC curve

FIGURE 2. Criteria for steganalysis

passive judging from the nature of her work on examining the media in transmission. If she makes $\boldsymbol{Y}' \neq \boldsymbol{Y}$ in order to foil all possible covert communications between Alice and Bob, she is called an active warden. If she only takes actions when $\boldsymbol{Y}$ is found suspicious, she is a passive warden. In the passive warden case, which is the main focus of this paper, once Wendy can differentiate $\boldsymbol{Y}$ from $\boldsymbol{X}$, the steganographic method is considered broken. Note that this model only aims to explain the concepts of steganography and steganalysis, but not to detail the way on how to conduct the practice.

2.2. **Evaluation Criteria.** In order to reasonably evaluate the performance of various kinds of steganographic and steganalytic methods, it is necessary to define some criteria acceptable to the majority. Moreover, the evaluation criteria may also lead us to the right direction to improve the techniques.

2.2.1. *Criteria for Steganography.* Three common requirements, security, capacity, and imperceptibility, may be used to rate the performance of steganographic techniques.

Security. Steganography may suffer from many active or passive attacks, correspondingly in the prisoner's problem when Wendy acts as an active or passive warden. If the existence of the secret message can only be estimated with a probability not higher than random guessing in the presence of some steganalytic systems, steganography may be considered secure under such steganalytic systems. Otherwise we may claim it to be insecure. The definition of security is further discussed in Section 2.3.

Capacity. To be useful in conveying secret message, the hiding capacity provided by steganography should be as high as possible, which may be given in absolute measurement (such as the size of secret message), or in relative value (called data embedding rate, such as bits per pixel, bits per non-zero discrete cosine transform coefficient, or the ratio of the secret message to the cover medium, etc.).

Imperceptibility. Stego images should not have severe visual artifacts. Under the same level of security and capacity, the higher the fidelity of the stego image, the better. If the resultant stego image appears innocuous enough, one can believe this requirement to be satisfied well for the warden not having the original cover image to compare.

2.2.2. *Criteria for Steganalysis.* The main goal of steganalysis is to identify whether or not a suspected medium is embedded with secret data, in other words, to determine the testing medium belong to the cover class or the stego class. If a certain steganalytic

method is used to steganalyze a suspicious medium, there are four possible resultant situations.

- True positive (TP), meaning that a stego medium is correctly classified as stego.
- False negative (FN), meaning that a stego medium is wrongly classified as cover.
- True negative (TN), meaning that a cover medium is correctly classified as cover.
- False positive (FP), meaning that a cover medium is wrongly classified as stego.

Confusion Matrix. When applying a steganalytic method on a testing data set, which may consist of cover and stego media, a 2×2 confusion matrix[8], which is illustrated in Figure 2(a), can be constructed, representing the dispositions of the instances in the set. Based on this matrix, some evaluation metrics can be defined.

$$\text{TP Rate} = \frac{\text{TPs}}{\text{TPs + FNs}},$$
$$\text{FP Rate} = \frac{\text{FPs}}{\text{TNs + FPs}},$$
$$\text{Accuracy} = \frac{\text{TPs + TNs}}{\text{TPs + FNs +TNs + FPs}},$$
$$\text{Precision} = \frac{\text{TPs}}{\text{TPs + FPs}}.$$

Receiver Operating Characteristic (ROC) Curve. The performance of a steganalytic classifier may be visualized by an ROC curve [8], in which true positive rate is plotted on the vertical axis and false positive rate is plotted on the horizontal axis (see Figure 2(b)). If the area under the ROC curve (AUC) is larger, the performance of the steganalytic method is better. For example, it can be observed from Figure 2(b) that the performance of ROC curve C is better than B, and B is better than A.

2.3. **Steganographic Security.** Security is the most important evaluation criterion in steganography and steganalysis. There are several kinds of definition of steganographic security, each of which are defined from different viewing angles.

2.3.1. *Information Theoretical Security.* From the point of view of information theory, Cachin [9] quantified the security of a steganographic system in terms of the relative entropy between the distribution of $\boldsymbol{X}$, denoted by $P_{\boldsymbol{X}}$, and that of $\boldsymbol{Y}$, denoted by $P_{\boldsymbol{Y}}$, in face of passive attacks. The relative entropy between $P_{\boldsymbol{X}}$ and $P_{\boldsymbol{Y}}$ is defined as [9]

$$D\left(P_{\boldsymbol{X}} || P_{\boldsymbol{Y}}\right) = E_{P_{\boldsymbol{X}}} \log \frac{P_{\boldsymbol{X}}}{P_{\boldsymbol{Y}}} \tag{1}$$

Based on this definition, if $D\left(P_{\boldsymbol{X}} || P_{\boldsymbol{Y}}\right) \leq \varepsilon$, the steganographic system is said to be ε-secure under passive attack. If $\varepsilon = 0$, the steganographic technique is called perfectly secure.

2.3.2. *ROC-based Security.* In ref. [6], several shortcomings in the information theoretical definition of steganographic security are discussed and an alternative security measure based on steganalyzer's ROC performance is then proposed. As stated in Section 2.2.2, ROC is a plot of false positive rate versus true positive rate, which represents the achievable performance of a steganalytic system. Therefore, the steganographic security under practical steganalyzers may be defined as the following.

- A steganographic technique is said to be γ-secure *with respect to (w.r.t.)* a steganalyzer if $|\text{TP Rate} - \text{FP Rate}| \leq \gamma$, where $0 \leq \gamma \leq 1$.
- A steganographic technique is said to be perfectly secure *w.r.t.* a steganalyzer if $\gamma = 0$.

2.3.3. *Maximum Mean Discrepancy Security.* Steganalytic methods often map images into a feature space, in which cover images and stego images may have different distributions since they can be considered as samples generated from two different sources. Maximum mean discrepancy (MMD)[10], a statistical method for testing if two kinds of samples are generated from the same distribution, may be suitable for benchmarking steganography since it is numerically stable even in high-dimensional space. It is defined as

$$\mathrm{MMD}[\mathcal{F}, \boldsymbol{X}, \boldsymbol{Y}] = \sup_{f \in \mathcal{F}} \left(\frac{1}{D} \sum_{i=1}^{D} f(x_i) - \frac{1}{D} \sum_{i=1}^{D} f(y_i) \right) \tag{2}$$

where $\boldsymbol{X} = \{x_1, \ldots, x_D\}, \boldsymbol{Y} = \{y_1, \ldots, y_D\}$ are the samples from $P_{\boldsymbol{X}}$ and $P_{\boldsymbol{Y}}$, respectively. $\mathcal{F}$ is a class of function which should be chosen carefully. More details on selecting the function $\mathcal{F}$ are covered in ref. [10].

3. **Image Steganography.** Although steganography for binary images [11, 12] and 3-D images [13] have some progresses, researches mainly concentrate on hiding data in gray-scale images and color images. Since the luminance component of a color image is equivalent to a gray-scale image, we focus on the steganography for gray-scale images. Besides, it is generally considered that gray-scale images are more suitable than color images for hiding data [14] because the disturbance of correlations between color components may easily reveal the trace of embedding. If not specified, the images in this paper are referred to 8-bit gray-scale images. Owing to the fact that bitmap/raw and JPEG images are of great interests in steganography community, we focus on spatial steganography and JPEG steganography. Moreover, since the data extraction steps are usually the inverse operations of the data embedding steps, we mainly describe the data embedding approaches for each steganographic method in the following sub-sections.

3.1. **Spatial steganography.** The common ground of spatial steganography is to directly change the image pixel values for hiding data. The embedding rate is often measured in bit per pixel (bpp). According to the embedding manner, we review six major kinds of steganography in the following.

3.1.1. *Least Significant Bit (LSB) Based Steganography.* LSB based steganography is one of the conventional techniques capable of hiding large secret message in a cover image without introducing many perceptible distortions[15]. It works by replacing the LSBs of randomly selected pixels in the cover image with the secret message bits. The selection of pixels may be determined by a secret key. The embedding operation of LSB steganography may be described by the following equation

$$y_i = 2\lfloor \frac{x_i}{2} \rfloor + m_i, \tag{3}$$

where m_i, x_i, and y_i are the i-th message bit, the i-th selected pixel value before embedding, and that after embedding, respectively. Many steganographic tools using the LSB based steganographic technique, such as Steghide, S-tools, Steganos, etc, are available on the Internet[1].

Let $\{P_{\boldsymbol{X}}(x=0), P_{\boldsymbol{X}}(x=1)\}$ denote the distribution of the least significant bits of a cover image, and $\{P_{\boldsymbol{m}}(m=0), P_{\boldsymbol{m}}(m=1)\}$ denote the distribution of the secret binary message bits. Generally, in order to protect the secrecy, the to-be-hidden message may be compressed or encrypted before being embedded. Hence, the distribution of message may be assumed to approximate a uniform distribution, that is, $\{P_{\boldsymbol{m}}(m=0) \simeq P_{\boldsymbol{m}}(m=$

[1] `http://www.stegoarchive.com`

$1) \simeq 1/2\}$. Besides, the cover image and message may also be assumed to be independent. Thus, the noise introduced to the image (thereafter stego-noise) may be modeled as

$$P_{+1} = \frac{p}{2} P_{\boldsymbol{X}}(x = 0),\ P_0 = 1 - \frac{p}{2},\ P_{-1} = \frac{p}{2} P_{\boldsymbol{X}}(x = 1), \tag{4}$$

where p is the embedding rate in bpp.

From the embedding operation described above, it is easy to know that the secret message bits may be extracted directly from the LSBs of these pixels which are selected during embedding.

3.1.2. *Multiple Bit-planes Based Steganography.* The methodology of LSB embedding can be easily extended to hiding data in multiple bit-planes. But one major defect of this kind of extension is that the non-adaptive embedding manner may reduce the perceptual quality of a stego image if some high bit-planes are involved in embedding arbitrarily without considering the local property. To address this problem, Kawaguchi and Eason[16] developed the bit-plane complexity segmentation (BPCS) steganography. In this method, the raw image which is represented in pure-binary coding (PBC) system will be firstly converted to canonical Gray coding (CGC) system. Then the image is decomposed to a set of binary images according to the bit-plane. Next, for each candidate embedding CGC bit-plane, its corresponding binary image is divided into consecutive and non-overlapping blocks of size $2^L \times 2^L$, where $L = 3$ is a recommended choice. If the complexity of the image-block, computed by

$$\alpha = \frac{k}{2 \times 2^L \times (2^L - 1)}, \tag{5}$$

is larger than a predefined threshold α_0, such a block is regarded as noise-like and suitable for data embedding. The k in Eq. (5) stands for the total number of black-and-white borders in the block. At the same time, secret data are grouped into a series of data-blocks with the size $2^L \times 2^L$ and their complexities are also computed by eq. (5). If the complexity of a data-block is less than α_0, such a block is processed by a conjugation operation[16] and the complexity of the conjugated data-block will be $(1-\alpha)$, larger than α_0. Then the noise-like data-blocks will replace the noise-like image-blocks to carry data. And the whole image after data embedding is transformed back to PBC system. The embedding rate of BPCS steganography may achieve as high as 4 bpp without causing severe visual artifacts.

3.1.3. *Noise-adding Based Steganography.* The embedding effect of "pairs of value" (PoV) exists in LSB steganography and may lead to successful steganalysis[17] (see Section 4.1.1 for details). In order to avoid PoV statistical attack, LSB matching[18, 19, 20], which is a minor modification of LSB steganography, is proposed. Instead of replacing the LSBs of the cover image pixels, LSB matching adds or subtracts them by 1 if they does not match the message bits.

In fact, LSB matching may be considered as a special case of $\pm k$ steganography[21] with $k = 1$, which increases or decreases the pixel value by k to match its LSB with the binary message bit. The distortion due to non-adaptive $\pm k$ embedding may be modeled as an additive independent identically distributed (*i.i.d.*) noise signal with the following probability mass function (PMF)

$$P_{+k} = \frac{p}{4},\ P_0 = 1 - \frac{p}{2},\ P_{-k} = \frac{p}{4}, \tag{6}$$

where p is the embedding rate in bpp.

Fridrich[22] presented another novel noise-adding steganography, known as stochastic modulation steganography. Message bits are embedded in the cover image by adding a

weak noise signal with a specified but arbitrary probabilistic distribution. A high hiding capacity (about 0.8 bpp) may be achieved with the use of a well-designed parametric parity function. The parametric parity function $p(x, z)$ used in stochastic modulation steganography is required to satisfy the anti-symmetric property for x, i.e. $p(x + z, z) = -p(x - z, z)$ $(z \neq 0)$. The definition of the parity function proposed in ref. [22] is given as follows.

- If $x \in [1, 2z]$, $p(x, z) = \begin{cases} (-1)^{x+z} & \text{if } z > 0, \\ 0 & \text{if } z = 0. \end{cases}$
- If $x \notin [1, 2z]$, $p(x, z)$ can be computed according to the anti-symmetric property.

The embedding procedure of stochastic modulation can be described as follows. Firstly, sequential or random visiting path and the to-be-added stego-noise ξ_n are generated using a secret key. Then for the pixel x_i along the visiting path, one sample n_i of the stego-noise ξ_n is rounded off to an integer z_i. If $z_i = 0$, the pixel x_i is skipped and at the same time the next stego-noise sample is input and rounded. If $z_i \neq 0$, the pixel x_i will be modified according to the value of the parity function. That is,

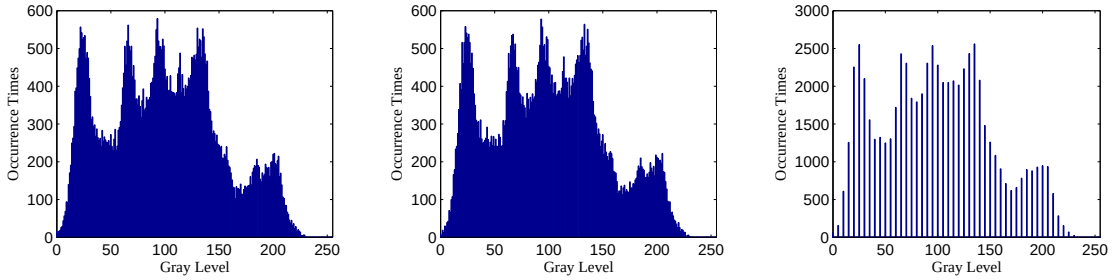
$$\begin{aligned} \text{if} \quad & p(x_i + z_i, z_i) = m_k \quad \text{then} \quad y_i = x_i + z_i, \\ \text{elseif} \quad & p(x_i + z_i, z) = -m_k \quad \text{then} \quad y_i = x_i - z_i. \end{aligned} \tag{7}$$

where m_k is the k-th message bit. During the embedding process, those pixels out of the range of [0, 255] will be truncated to the nearest values in this range with the desired parity.

Though the embedding operations of LSB matching and $\pm k$ steganography are different from that of LSB steganography, their methods of extracting the secret message bits are the same as the one stated in Section 3.1.1. For message extraction in stochastic modulation steganography, the same rounded stego-noise sequence z_i is generated from the stego key as is done during message embedding, follow the same pseudo-random path in the stego image, and apply the parity function $p(x, z)$ to the pixel values. The non-zero parity values form the secret message.

3.1.4. *Prediction Error Based Steganography.* In order to maintain image visual quality, it is intuitive to think that secret data should be hidden in complex areas of the image. To evaluate the local complexity, one way is to use the pixel prediction error. The large the prediction error, the more obvious the local fluctuation. Data can be hidden into the prediction errors. Using a pixel's neighboring pixel is a simple way to predict the current pixel value and thus their difference can be considered as a kind of prediction error. In the pixel value differencing (PVD) steganography[23], an image is partitioned into non-overlapping and consecutive groups of two neighboring pixels. The to-be-embedded secret data are hidden into the difference values. Suppose two neighboring pixels, x_i and x_{i+1}, are used and their difference value is $d_i = x_{i+1} - x_i$, where $0 \leq |d_i| \leq 255$. A large $|d_i|$ means a complex block. Then classify $|d_i|$ into a set of contiguous ranges, denoted by R_k, where $k = 0, 1, ..., K - 1$ is the range index. Denote l_k, u_k, and w_k as the lower bound, the upper bound, and the width of R_k, respectively. The value of w_k is designed to be a power of 2. If $|d_i| \in R_k$, the corresponding two pixels are expected to carry $\log_2(w_k)$ bits. That is, their pixel values are changed so that the absolute value of their new difference equals to $|d'_i| = |y_{i+1} - y_i| = l_k + b_i$, where b_i is the decimal value of the to-be-embedded bits. The embedding operation can be described as

$$(y_i, y_{i+1}) = \begin{cases} (x_i - r_c, x_{i+1} + r_f) & \text{if } d_i \text{ is odd}, \\ (x_i - r_f, x_{i+1} + r_c) & \text{if } d_i \text{ is even}, \end{cases} \tag{8}$$



(a) Cover (b) Stego ($\Delta = 2$) (c) Stego ($\Delta = 10$)

FIGURE 3. The histogram of a cover image (a), the histogram of of a stego image with $\Delta = 2$ (b), and that of a stego image with $\Delta = 10$ (c).

where $r_c = \lceil \frac{d'_i - d_i}{2} \rceil$ and $r_f = \lfloor \frac{d'_i - d_i}{2} \rfloor$. In this way, the embedding distortion is distributed almost equally in two pixels. In Bob's side, the difference values can be obtained. If $|d'_i| \in R_k$, the decimal value of the embedded bits is computed as $b_i = |d'_i| - l_k$.

3.1.5. *Modulo Operation Based Steganography.* In multiple base notational system (MBNS) steganography [24], binary secret data are converted into symbols represented in a notational system with variable bases. The conversion can be done by using simple arithmetic as described in ref. [24]. The pixel value is modified to

$$y_i = \underset{v \in [0,255],\ \mathrm{mod}\ (v,b_i)=d_i}{\arg\min} |v - x_i|, \tag{9}$$

where b_i and d_i are the base value and the corresponding symbol for the to-be-modified pixel x_i, respectively. The $\mathrm{mod}(v, b_i)$ is an modulo operation which computes the remainder of division of v by b_i. In this way, the remainder of division of new pixel value y_i by the base value b_i equals to the to-be-embedded symbol d_i. The base value is determined by image local property and the pixel y_i can carry $\log_2(b_i)$ bits secret data. The larger the local variation, the larger the base is, and the more information bits can be hidden in the pixel. MBNS steganography takes advantage of the human visual system. Its embedding rate may even achieve 2 bpp for some images. In the data extraction side, the base value b_i can be retrieved from the image and thus $d_i = \mathrm{mod}(y_i, b_i)$. Then the symbols are transformed back to binary data. Since the data embedding and data extraction processes are based on a modulo operation, we regard such type of steganography as modulo operation based steganography.

3.1.6. *Quantization Based Steganography.* Quantization index modulation (QIM)[25] is a commonly used data embedding technique in digital watermarking and it can be employed for steganography. It quantizes the input signal x to the output y with a set of quantizers, i.e., $\boldsymbol{Q}_m(\cdot)$. Using which quantizer for quantization is determined by the message bit m. A standard scalar QIM with quantization step Δ for embedding binary data can be simply described as:

$$y_i = \boldsymbol{Q}_m(x_i) = \begin{cases} \Delta \left\lfloor \frac{x_i}{\Delta} + \frac{1}{2} \right\rfloor & \text{if } m_i = 0, \\ \Delta \left\lfloor \frac{x_i}{\Delta} \right\rfloor + \frac{\Delta}{2} & \text{if } m_i = 1. \end{cases} \tag{10}$$

As explained in ref. [3] and illustrated in Figure 3, if the standard QIM is employed to spatial domain, the histogram will show a sign of discreteness in the integer multiple of $\Delta/2$, especially when $\Delta > 2$. It is unusual for a spatial image to have such a kind of quantization phenomenon. Therefore QIM is often employed to the coefficients in the

transform domain which are needed to be quantized. For example, QIM can be used with JPEG compression, such as the method described in ref. [26].

A variant of QIM is called dither modulation (DM)[25, 27]. Unlike QIM which produces the output values only at the reconstruction points of quantizers, DM can produce the output signal covering all of the values of the input signal. Such capability is achieved by adding a dither signal to the input signal before quantization and subtracting it after quantization. That is,

$$y_i = \boldsymbol{Q}_m(x_i + d_i) - d_i. \tag{11}$$

The dither signal d_i is determined by a key and uniformly distributed over $[-\Delta/4, \Delta/4)$. DM can be applied to spatial image to avoid making the histogram sparse, but it is also more often used for transform coefficients.

3.2. **JPEG steganography.** JPEG is the common format of the images produced by digital cameras, scanners, and other photographic image capture devices. Therefore, hiding secret information into JPEG images may provide better camouflage. Most of the steganographic schemes embed data into the non-zero alternate current (AC) discrete cosine transform (DCT) coefficients of JPEG images. As a result, the embedding rate of JPEG steganographic is often evaluated in bit per non-zero AC DCT coefficient (bpac). We review five major JPEG steganographic methods in the following.

3.2.1. *JSteg/JPHide.* Jsteg [28] and JPHide [29] are two classical JPEG steganographic tools utilizing the LSB embedding technique. JSteg embeds secret information into a cover image by successively replacing the LSBs of non-zero quantized DCT coefficients with secret message bits. Unlike JSteg, the quantized DCT coefficients that will be used to hide secret message bits in JPHide are selected at random by a pseudo-random number generator, which may be controlled by a key. Moreover, JPHide modifies not only the LSBs of the selected coefficients, it can also switch to a mode where the bits of the second least significant bit-plane are modified.

3.2.2. *F5.* F5 steganographic algorithm was introduced by Westfeld[30]. Instead of replacing the LSBs of quantized DCT coefficients with the message bits, the absolute value of the coefficient is decreased by one if it is needed to be modified. The author argued that this type of embedding cannot be detected using the chi-square attack[17]. The F5 algorithm embeds message bits into randomly-chosen DCT coefficients and employs matrix embedding that minimizes the necessary number of changes to hide a message of certain length. In the embedding process, the message length and the number of non-zero AC coefficients are used to determine the best matrix embedding that minimizes the number of modifications of the cover image.

3.2.3. *OutGuess.* OutGuess[31] is provided by Provos as UNIX source code. There are two famous released versions: OutGuess-0.13b, which is vulnerable to statistical analysis, and OutGuess-0.2, which includes the ability to preserve statistical properties. When we talk about the OutGuess, it is referred to OutGuess-0.2. The embedding process of OutGuess is divided into two stages. Firstly, OutGuess embeds secret message bits along a random walk into the LSBs of the quantized DCT coefficients while skipping 0's and 1's. After embedding, corrections are then made to the coefficients, which are not selected during embedding, to make the global DCT histogram of the stego image match that of the cover image. OutGuess cannot be detected by chi-square attack[17].

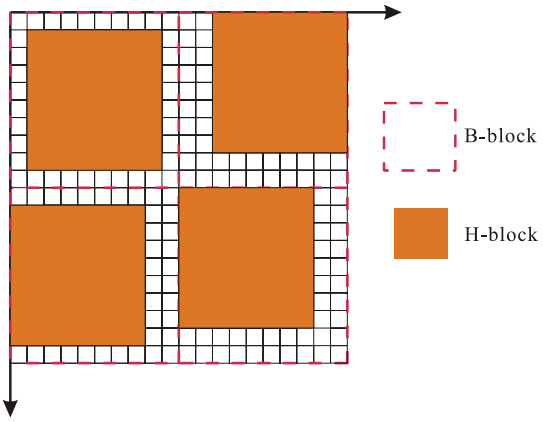
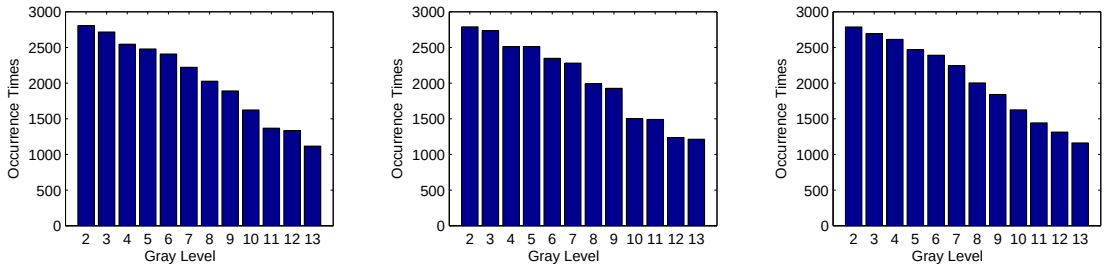


FIGURE 4. Illustration of B-blocks and H-blocks in YASS

3.2.4. *MB.* Sallee[32] presented a general framework for performing steganography and steganalysis using a statistical model of the cover media. The proposed example steganographic method for JPEG images, named model-based steganography (MB), achieves a high message capacity while remaining secure against several first order statistical attacks. MB adapts the division of the carrier into a deterministic random variable $\mathbf{X}_{det}$ and an in-deterministic one $\mathbf{X}_{indet}$. And a suitable model is employed to describe the distribution of $\mathbf{X}_{indet}$, which reflects the dependencies with $\mathbf{X}_{det}$. The general model is parameterized with the actual values of $\mathbf{X}_{det}$ of a concrete cover image, which leads to a cover specific model. The purpose of this model is to determine the conditional distributions $P(\mathbf{X}_{indet}|\mathbf{X}_{det} = x_{det})$. Then, an arithmetic decompression function is used to fit uniformly distributed message bits to the required distribution of $\mathbf{X}_{indet}$, thus replacing $\mathbf{X}_{indet}$ by $\mathbf{X}^*_{indet}$, which has similar statistic properties and contains the confidential message.

3.2.5. *YASS.* Yet Another Steganographic Scheme (YASS) [33] belongs to JPEG steganography but it does not embed data in JPEG DCT coefficients directly. Instead, an input image in spatial representation is firstly divided into blocks with a fixed large size, and such blocks are called big blocks (or B-blocks). Then within each B-block, an 8×8 sub-block, referred to as embedding host block (or H-block), is randomly selected with a secret key for performing DCT. The B-blocks and H-blocks are illustrated in Figure 4. Next, secret data encoded by error correction codes are embedded in the DCT coefficients of the H-blocks by QIM. Finally, after performing the inverse DCT to the H-blocks, the whole image is compressed and distributed as a JPEG image. For data extraction, image is firstly JPEG-decompressed to spatial domain. Then data are retrieved from the DCT coefficients of the H-blocks. Since the location of the H-blocks may not overlap with the JPEG 8×8 grid, the embedding artifacts caused by YASS are not directly reflected in the JPEG DCT coefficients. The self-calibration process [34, 35], a powerful technique in JPEG steganalysis for estimating the cover image statistics, is disabled by YASS. Another advantage of YASS is that the embedded data may survive in the active warden scenario. Recently Yu et al [36] proposed a YASS-like scheme to enhance the security performance of YASS via enhancing block randomization. The comparative security performance of YASS, F5 and MB against state-of-the-art steganalytic methods can be found in recent work of Huang et al [37].

4. **Image Steganalysis.** Steganalysis can be regarded as a two-class pattern classification problem which aims to determine whether a testing medium is a cover medium or



(a) Original histogram (b) LSB steganography (c) LSB Matching

FIGURE 5. The histogram of a sample cover image (a), that of the stego images produced by LSB steganography (b), and that of the stego images produced by LSB matching steganography(c), respectively.

a stego one. According to its application fields, it can be divided into specific methods and universal methods. A specific steganalytic method fully utilizes the knowledge of a targeted steganographic technique and may only be applicable to such a kind of steganography. A universal steganalytic method can be used to detect several kinds of steganography. Usually universal methods do not require the knowledge of the details of the embedding operations. Therefore, it is also called blind method. Some methods can be considered as "semi-universal". For example, the methods in ref. [34, 35, 38, 39] can reliably detect many JPEG steganographic schemes but may not be effective to spatial steganography. We still regard these methods in the universal category.

4.1. **Specific Approaches.** A specific steganalytic method often takes advantage of the insecure aspect of a steganographic algorithm. We present some specific steganalytic methods for attacking the steganographic schemes introduced in Section 3.

4.1.1. *Attacking LSB steganography.* As mentioned previously, LSB steganography was put into use in many steganographic tools very early and has been one of the most important spatial steganographic techniques. Accordingly, much work has been done on steganalyzing LSB steganography in the initial stage of the development of steganalysis. And many steganalytic methods toward LSB steganography have been proved most successful, such as Chi-square (χ^2) statistical attack [17, 40], RS analysis [14], sample pair analysis (SPA) analysis [41], weighted stego (WS) analysis [42], and structural steganalysis [43, 44], etc.

As regards LSB steganography, some of the LSBs of a cover image will be flipped when they differs from the message bits, which is discussed in details in Section 3.1.1. Without loss of generality, the message bits may be considered to be uniformly distributed, which is usually the case when they are compressed or encrypted ahead of embedding. Then the flipping $2n \leftrightarrow 2n+1$ ($n = 0, 1, \cdots, 127$ for a gray-scale image) may result in the occurrence times of both values of each PoV $(2n, 2n+1)$, denoted by O_{2n} and O_{2n+1} respectively, becoming more equal than those of the original cover image, which can be seen from Figure 5(b) and 5(a). The more uniformly message bits are hidden into the cover image, the more the occurrence times of a PoV will be equal. But the sum of their occurrence times $O_{2n} + O_{2n+1}$ stays the same. Thus, the arithmetic mean of the sum, denoted by $O_e = \frac{O_{2n}+O_{2n+1}}{2}$ may be taken as the theoretically expected frequency in the Chi-square test for the frequency of occurrence of $2n$ or $2n+1$ [17]. Then the χ^2 statistic may be given as $\chi^2_{k-1} = \sum \frac{(O_{2n}-O_e)^2}{O_e}$ with $k-1$ degrees of freedom. And the probability

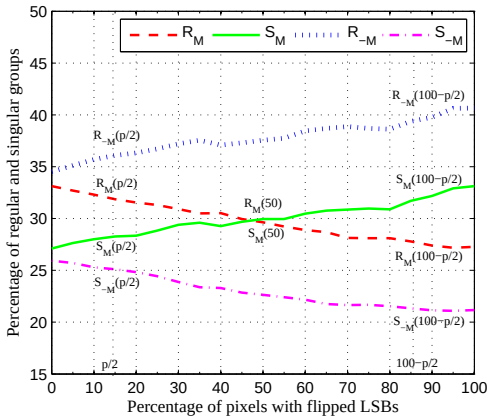


FIGURE 6. RS diagram of a gray-scale 128 × 128 Lena. The *x-axis* is the percentage of pixels with flipped LSBs, the *y-axis* is the relative number of regular and singular groups with masks M and $-M$, $M = [0\ 1\ 1\ 0]$.

of embedding p can be calculated by

$$p = 1 - \frac{1}{2^{\frac{k-1}{2}}\Gamma(\frac{k-1}{2})} \int_0^{\chi^2_{k-1}} e^{-\frac{x}{2}} x^{\frac{k-1}{2}-1} dx \tag{12}$$

where Γ is the Euler Gamma function.

The quantitative analysis of LSB steganography was firstly addressed by Fridrich et al. [14] Their method is well known as RS analysis. By defining a discrimination function f, which maps a group of n neighboring pixels $(x_1, x_2, \cdots, x_n)$ into a real number, and an invertible flipping operation F, a pixel group G is classified into one of the three types: R,S, and U.

$$\begin{aligned} \text{Regular groups:} \quad & G \in R \Longleftrightarrow f(F(G)) > f(G) \\ \text{Singular groups:} \quad & G \in S \Longleftrightarrow f(F(G)) < f(G) \\ \text{Unusable groups:} \quad & G \in U \Longleftrightarrow f(F(G)) = f(G) \end{aligned} \tag{13}$$

where $F(G)$ means apply the operation F on each element of G. Different flipping may be conducted on different pixels and the assignment of flipping to each pixel in G is given by a mask M. Fridrich et al. observed that the approximate equality existing between the number of regular (singular) groups for mask M denoted by R_M (S_M) and that for mask $-M$ denoted by R_{-M} (S_{-M}) may be destroyed by LSB steganography to a corresponding degree with the length of the message bits, which is well illustrated by the RS diagram in Figure 6, where the R_{-M} and S_{-M} curves can be well modeled with a straight line, and the "inner" curves R_M and S_M follow a parabola. Then the message length may be calculated with these models and the special points in the RS diagram.

Many other steganalytic techniques [41, 42, 43, 44] have been proposed in recent years. The success of most of these methods is based on the fact that the pixel/coefficient values are changed within the PoV, i.e., $2n \leftrightarrow 2n+1$. Note that some steganalytic methods, for example, the Chi-square attack [17, 40], are effective to LSB steganography for spatial images as well as JPEG images. The fact that LSB steganography is vulnerable to attack implies that high imperceptivity does not guarantee a high security level.

4.1.2. *Attacking LSB Matching Steganography.* It may be seen from Figure 5(c) that the equal trend of the frequency of occurrence of PoVs no longer exists for LSB matching steganography. Thus many steganalytic methods toward LSB steganography turn out be invalid. LSB matching, or more general $\pm k$ steganography, may be modeled in the

context of additive noise independent of the cover image, which is discussed in Section 3.1.3.

The effect of additive noise steganography to the image histogram is equivalent to a convolution of the histogram of the cover image and the stego-noise PMF. It may be analyzed more conveniently in the frequency domain [45]. Let the histogram characteristic function (HCF) be the discrete Fourier transform (DFT) of a histogram. The histogram characteristic function center of mass (HCF-COM), which gives a general information about the energy distribution in HCF, is exploited to capture the low pass filter effect of the additive noise. The HCF-COM can successful detect the steganographic techniques of additive noise type.

In ref. [46], Ker's experimental results showed that the HCF-COM-based steganalytic method performed quite good for color images, but it turned out to have very poor performance for gray-scale images. Ker found that the reason lied in the high variability of the cover images' HCF. Therefore, a down-sampled image by a factor of two in both dimensions and processed by a straightforward averaging filter was employed to calibrate the HCF-COM of the full-sized image [46]. In view of the variation between the magnitudes of the HCF-COM of a cover image, denoted by $\mathcal{C}(H[k])$, and that of the down-sampled image, denoted by $\mathcal{C}(H'[k])$, the ratio $\mathcal{C}(H[k])/\mathcal{C}(H'[k])$ is then proposed as a dimensionless discriminator. Another way of applying the HCF-COM is also introduced by computing the adjacency histogram. The HCF-COM detector based on $\mathcal{C}(H[k])/\mathcal{C}(H'[k])$ and that based on the adjacency histogram are proved by extensive experimental that both of them produce reliable detectors for LSB matching steganography in gray-scale images. A novel calibration-based detectors calculated on the difference image to detect LSB matching is recently investigated in ref. [47]. By combining techniques of pixel selection and utilizing low-frequency DFT coefficients, the new detectors outperform the Ker's calibrated version and are capable of detecting LSB matching in gray-scale image even when the embedding rate is low, especially for compressed images.

Besides the steganalytic algorithms summarized above, there are still several other targeted methods of steganalyzing LSB matching [48, 49, 50]. Zhang et al. [48] observed that the local maxima of an image's gray-level or color histogram decrease and the local minima increase. Consequently, the sum of the absolute differences between the local extrema and their neighbors in the histogram of a cover image will be greater than that of the stego image. This property is then used to construct a new discriminant feature for steganalysis. Later, the algorithm of Zhang et al. was modified [49] to deal with border effects associated with the 1-D intensity histogram, and extended to include statistics associated the amplitude of local extrema in the 2-D adjacency histogram. In ref. [50], a new image is first produced by combining the least two significant bit-planes and is then divided into 3×3 overlapped sub-images. The sub-images are grouped into four types T_i $(i = 1, 2, 3, 4)$, where i is the number of gray levels in a sub-image. Via embedding a random sequence by LSB matching and then computing the alteration rate of the number of elements in T_1, the alteration rate is found to be higher in cover image than in the corresponding stego image. And this new finding is used as the discrimination rule for the detection of LSB matching.

4.1.3. *Attacking Stochastic Modulation Steganography.* It is reported in [51] that the horizontal pixel difference histogram of a natural image can be modeled as a generalized Gaussian distribution (GGD). However, as stated in 3.1.3, stochastic modulation steganography adds stego-noise with a specific probability distribution into the cover image to embed secret message bits. The embedding effect of adding stego-noise may disturb the distribution of the cover natural image. A quantitative approach to steganalyze stochastic

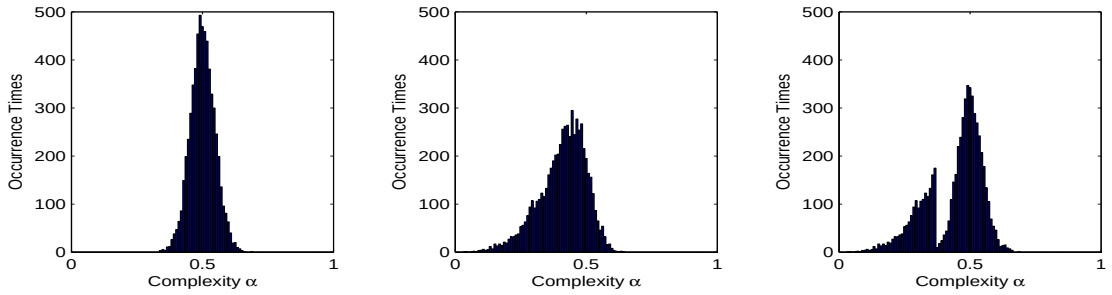


FIGURE 7. The complexity histogram of some data-blocks (a), that of the image-blocks of 5th most significant bit-plane of a cover image (b), and that of a stego image (with complexity threshold $\alpha_0 = 0.375$).

modulation steganography was presented in [52, 53]. For the non-adaptive stochastic modulation steganography, the stego-noise added during embedding may assumed to be independent from the cover image. The distribution of stego-image's pixel difference is thus approximately equal to the convolution of the probabilistic distribution of the rounded stego-noise difference and that of the cover image's pixel difference. And the variance of the stego-noise, denoted by σ_n, may be estimated with the use of grid searching and goodness-of-fit test. Then the length p (in bpp) of the embedded secret message may be estimated by

$$p = 1 - erf(1/(2\sqrt{2}\sigma_n)) \tag{14}$$

where $erf(x) = 2/\sqrt{\pi}\int_0^x e^{-t^2}dt$. It is necessary to mention that the proposed estimator is not so robust for the two relied assumptions may not hold so well, which was analyzed in [53].

4.1.4. *Attacking the BPCS Steganography.* In BPCS steganography, the binary patterns of data-blocks are random and it is observed that the complexities of the data-blocks follow a Gaussian distribution with the mean value at 0.5 [54]. For some high significant bit-planes (e.g., the most significant bit-plane to the 5th significant bit-plane) in a cover image, the binary patterns of the image blocks are not random and thus the complexities of the image blocks do not follow a Gaussian distribution. If a histogram of the complexities of the image blocks is constructed, it is expected that the complexity histogram of a high significant bit-plane of a cover image is in a non-Gaussian-like shape. For a stego image, since the image blocks whose complexities being larger than the threshold α_0 are replaced by data-blocks, the complexities larger than α_0 will be replaced by the complexities of the data-blocks. Therefore, the complexity histogram will also be changed in the portion where the complexity is larger than α_0. It is expected that this portion will have a Gaussian-like shape. Besides, a valley can be found in the complexity histogram at the complexity threshold α_0. As a result, the presence of BPCS steganography can be revealed by observing the complexity histogram of high significant bit-planes, as proposed by Niimi et al. [54]. Figure 7 illustrates the complexity histogram of data-blocks, the complexity histogram of the image-blocks of the 5th most significant bit-plane of a cover image, and that of its stego image, respectively.

4.1.5. *Attacking the Prediction Error Based Steganography.* If there is no special scheme to prevent Wendy retrieving the correct prediction values, it is quite easy for Wendy to detect the steganographic method which utilizes prediction errors for hiding data, such as PVD

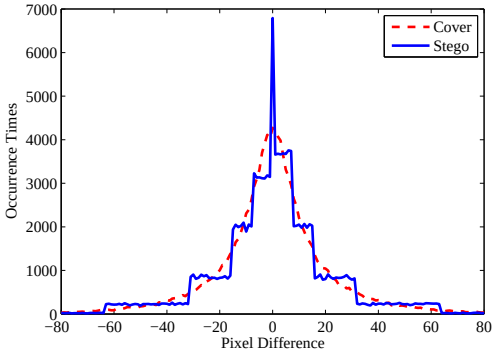
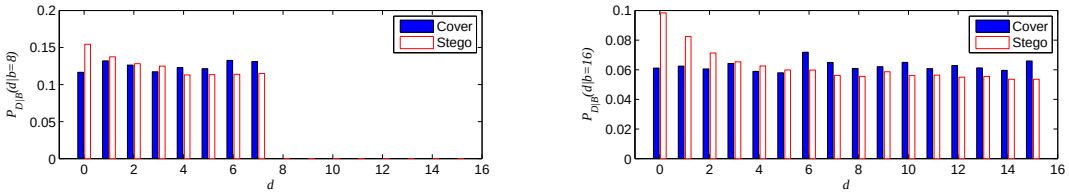


FIGURE 8. The histogram of pixel difference for a cover image and its stego image. The ranges in the stego image are set to as $R_1 = [0, 7]$, $R_2 = [8, 15]$, $R_3 = [16, 31]$, $R_4 = [32, 63]$, $R_5 = [64, 127]$, $R_6 = [128, 255]$



(a) $b = 8$ (b) $b = 16$

FIGURE 9. Conditional probability $P_{D|B}(d|b)$ for a cover-image and its stego-image

steganography. Zhang et al. [55] proposed a method for attacking PVD steganography based on observing the the histogram of the prediction errors. Since "0" and "1" are equally distributed in the binary secret data, the occurrence of the decimal values are also equally distributed in each R_k. The reason is very similar to the LSB steganography. Replacing the prediction errors with the secret data will make the histogram equalized in each R_k. Figure 8 shows the pixel value difference histogram of a cover image and that of a PVD stego image. It is quite easy to observe a "step effect" in the histogram and use the unusual phenomenon to launch an attack.

4.1.6. *Attacking the MBNS Steganography.* It's hard to observe any abnormality between a cover image and its MBNS stego image through the histogram of pixel values and the histogram of pixel prediction errors. In ref. [56], the authors observed and illustrated that given any base value, more small symbols are generated than large symbols in the process of converting binary data to symbols. Since the remainders of the division of pixel values by bases are equal to the symbols, the conditional probability $P_{D|B}$ can be used to discriminate the cover images and stego images, where B and D denote the random variable of the base and the remainder, respectively. For a given base, the following inequality holds in the stego image when the embedding rate is high.

$$P_{D|B}(D = 0|b) \geq P_{D|B}(D = 1|b) \geq \cdots \geq P_{D|B}(D = (b-1)|b) \tag{15}$$

Figure 9 shows the conditional probabilities when $b = 8$ and $b = 16$. To increase the robustness of the steganalytic method, it has been proposed to examine whether (16)

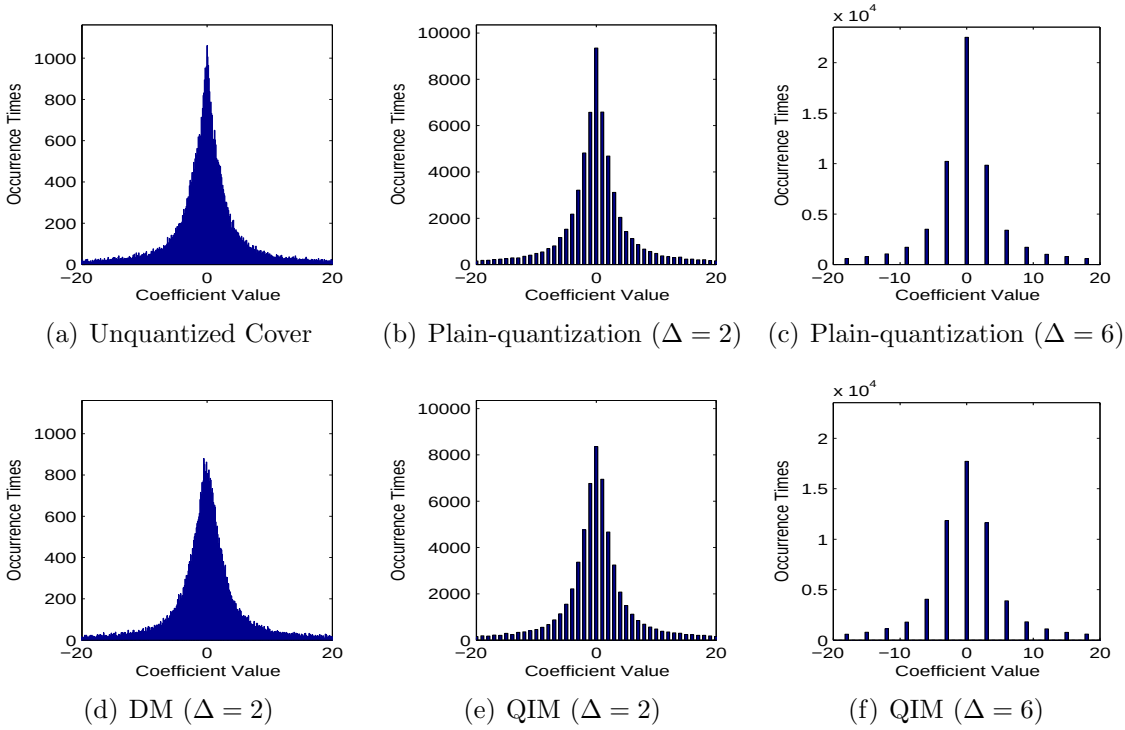
holds for the most frequently appeared base values in a testing image.

$$P_{D|B}(D=0|b) \geq \frac{1}{b-1}\sum_{i=1}^{b-1} P_{D|B}(D=i|b). \tag{16}$$

4.1.7. *Attacking QIM/DM.* The issue in steganalysis of QIM/DM has been formulated into two sub-issues by Sullivan et al. [57]. One is to distinguish the standard QIM stego objects from the plain-quantized (quantization without message embedding) cover objects. Another is to differentiate the DM stego objects from the unquantized cover objects. Figure 10 demonstrates the histogram of DCT coefficients of an unquantized cover image, plain-quantized images, QIM stego images, and DM stego image. Since the PMF of a QIM stego object, or the PDF (probability density function) of a DM stego object, has a relation with the PMF/PDF of its cover counterpart, if the PMF/PDF of the cover is known, a likelihood ratio test (LRT) can be conducted for optimal detection. It was noted in ref. [58] and confirmed by ref. [57] that if the PMF/PDF of the cover object follows a uniform distribution, it would be impossible to detect DM. In practice, the PMF/PDF of the coefficients in transform domain follows a Gaussian-like or Laplacian-like distribution, which means there is a large spike around the mean value. Therefore, it is possible to detect DM in real scenarios. For a Gaussian-like distribution, Sullivan et al. concluded that the detectability of QIM/DM is related to σ/Δ, where σ is a parameter measuring the concentration of PMF/PDF. Under the same σ, the larger the Δ, the easier the detection. This conclusion may be a bad news for Alice since she cannot have the robustness and the security at the same time. But Wendy cannot perform LRT in real life since she does not know the exact PMF/PDF of the cover object. Alternatively, a supervised learning scheme was practically employed in ref. [57] to use the PMF of the quantized coefficients as features for steganalyzing standard QIM. But the performance of steganalyzing DM has not been reported.

It was assumed the image coefficients are i.i.d. in Sullivan et al.'s work [57]. Malik et al. [59, 60, 61] proposed a series of methods which utilized the dependency among image coefficients when data are hiding in DCT coefficients by QIM/DM. In ref. [59], a random variable, named randomness mask and denoted by $R_{c_\mathbf{x}}$, has been defined to measure the similarity between the current DCT coefficient and the coefficients at the same frequency subband in the neighboring DCT blocks. Its value ranges from 0 to 1, where $R_{c_\mathbf{x}} = 0$ implies the maximum degree of similarity and $R_{c_\mathbf{x}} = 1$ indicates the minimum. Next, kernel density estimation is taken to estimate the density of the randomness mask. The density is then modeled by a Gamma density function. Finally, the skewness and the peak of the density function, are used for distinguishing between standard QIM stego images and quantized cover images via comparing the them with some predefined thresholds. This method has a high false positive rate when Δ is small. In an improved work in ref. [60], the detection performance is boosted.

The above mentioned methods [59, 60] can even be adapted to detect other JPEG steganographic schemes, which disturb the local correlation between coefficients. However, the performance of detecting the DM has not been reported. In ref. [61], two similar steganalytic schemes, both using approximate entropy (ApEn), had been proposed to detect QIM and DM, respectively. In the first scheme for detecting QIM, DCT coefficients in each individual AC DCT subband are firstly grouped into a sequence. Then the ApEn is calculated for each sequence. A high ApEn value indicates the degree of randomness of the sequence is high. It is observed that the ApEn values of the high frequency AC subband of a QIM stego image is always larger than that of a quantized cover image. This property is explored to detect the QIM stego image. However, such a method is still hard



(a) Unquantized Cover (b) Plain-quantization ($\Delta = 2$) (c) Plain-quantization ($\Delta = 6$)

(d) DM ($\Delta = 2$) (e) QIM ($\Delta = 2$) (f) QIM ($\Delta = 6$)

FIGURE 10. The histogram of DCT coefficients of a cover image without quantization (a), that of a cover image using plain-quantization with $\Delta = 2$ (b), that of a cover image using plain-quantization with $\Delta = 6$ (c), that of a DM stego image with $\Delta = 2$ (d), that of a QIM stego image with $\Delta = 6$ (e), and that of a QIM stego image with $\Delta = 6$ (f).

to distinguishing DM stego images and unquantized cover images. Hence the normalized ApEn (nApEn), obtained from dividing the ApEn by the variance of its corresponding DCT sequence, is used in the second scheme. To amplify the difference between cover and stego, a second testing image, named as DM re-embedded stego image, is generated by embedding some data into the testing image with DM. Then, the Euclidian distance between the nApEn of the testing image and that of its DM re-embedded stego image is computed. It is expected that the Euclidian distance of a stego image is smaller than that of a cover image. With a threshold, DM stego image and unquantized cover image may be differentiated.

4.1.8. *Attacking the F5 Algorithm.* Some crucial characteristics of the histogram of DCT coefficients, such as the monotonicity and the symmetry, are preserved by the F5 algorithm. But F5 does modify the shape of the histogram of DCT coefficients. This drawback is employed by Fridrich et al.[62] to launch an attack against F5. Let $h(d)$ be the total number of AC coefficients with absolute value equal to d in an image. In an F5 stego image, the first two values in the histogram ($d = 0$ and $d = 1$) experience the largest change during embedding. To facilitate the attack, a procedure of estimating the cover image's histogram from the stego image is taken in the steganalytic method as follows. Firstly, the stego image is decompressed to the spatial domain, then cropped by 4 columns, and re-compressed using the same quantization parameters as that of the original stego image. A blurring operation is applied as a preprocessing step to remove possible JPEG blocking artifacts from the cropped image before re-compressing. The resulting DCT coefficients

will provide the estimation of the cover image histogram. Then the probability of a non-zero AC coefficient being modified, denoted by β, may be estimated by the least square approximation minimizing the square error between the stego image histograms $h(0)$, $h(1)$ and those expected values obtained in the previous estimation procedure.

4.1.9. *Attacking OutGuess.* OutGuess preserves the shape of the histogram of DCT coefficients and thus it may not be easy to employ a quantitative steganalyzer to attack OutGuess with the statistics of DCT coefficients as that in attacking F5. Fridrich et al. [63] found a new path to detect OutGuess quantitatively by measuring the discontinuity along the boundaries of 8×8 JPEG grid. A spatial statistical feature, named blockiness, for an image is defined as

$$B = \sum_{i=1}^{\lfloor (M-1)/8 \rfloor} \sum_{j=1}^{N} |x_{8i,j} - x_{8i+1,j}| + \sum_{j=1}^{\lfloor (N-1)/8 \rfloor} \sum_{i=1}^{M} |x_{i,8j} - x_{i,8j+1}| \tag{17}$$

where $x_{i,j}$ is the gray level of the pixel at location (i, j) in an $M \times N$ image. It is observed that the blockiness linearly increases with the number of altered DCT coefficients. Suppose that some data are embedded into an input image. If the input image is innocent, the change rate of the blockiness between the input image and the embedded one will be large. If the input image already contains some data, the change rate will be smaller. The change rate of the blockiness can be used to estimate the embedding rate. In the steganalytic process, four corresponding images are generated from an input testing image. Denote the input image by T. The first image is generated by using OutGuess to embed data with maximal length to T and it is denoted as $\hat{S^0}$. The second one is created by decompressing the input image, cropping 4 columns, and then compressing the cropped image into JPEG with the same compression parameters as T. This image can approximate a cover image, and it is denoted as $\hat{C}$. The third image is formed through embedding data with maximal length to $\hat{C}$ and it is denoted as $\hat{S^1}$. The fourth one is generated by embedding some different data with maximal length to$\hat{S^1}$ and it is denoted as $\hat{S^2}$. $\hat{S^1}$ can simulate a stego image and $\hat{S^2}$ a twice data embedded stego image. The estimated embedding rate can be calculated as

$$p = \frac{[B(\hat{S^1}) - B(\hat{C})] - [B(\hat{S^0}) - B(T)]}{[B(\hat{S^1}) - B(\hat{C})] - [B(\hat{S^2}) - B(\hat{S^1})]} \tag{18}$$

where $B(T)$, $B(\hat{S^0})$, $B(\hat{C})$, $B(\hat{S^1})$, and $B(\hat{S^2})$ are the blockiness of T, $\hat{S^0}$, $\hat{C}$, $\hat{S^1}$, and $\hat{S^2}$, respectively.

4.1.10. *Attacking MB.* MB steganography uses a generalized Cauchy distribution model to control the data embedding operation. Therefore, the histogram of the DCT coefficients will fit the generalized Cauchy distribution well in a stego image. Bohme and Westfeld [64] observed that the histogram of the DCT coefficients in a natural image is not always conforming the distribution. There exist more outlier high precision bins in the histogram in a cover image than in a stego image. Judging from the number of outlier bins, cover images and stego images can be differentiated.

4.1.11. *Attacking YASS.* The locations of the H-blocks of YASS are determined by a key, which is not available to Wendy. Therefore, it may not be straightforward for Wendy to observe the embedding artifacts. Li et al. [65] observed that the locations of the H-blocks are not randomized enough in YASS. Specifically, the H-blocks are constrained to reside inside B-blocks. Define the origin of an block is the upper-left element in such a block. Along the main diagonal direction of a B-block, the first $(B - 7)$ elements are possible to be the origin of the H-block and the remaining 7 elements are definitely impossible to

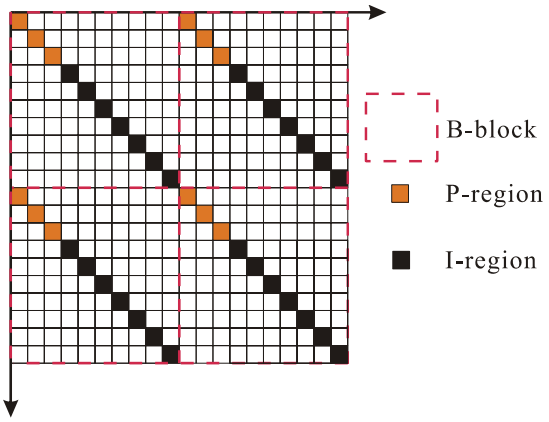


FIGURE 11. The P-regions and I-regions in 10×10 B-blocks

be the origin of the H-block. For simplicity we refer to these two kinds of element region in a B-block as P-region and I-region, respectively. Figure 11 shows the P-regions and I-regions in 10×10 B-blocks. These two kinds of regions bear different characteristics. Use a JPEG quantizer to quantize the 8×8 blocks whose origins are on the main diagonal direction of the B-blocks. It can be observed that more zero quantized coefficients are generated from P-region than I-region in a stego image. And a cover image does not show such a phenomenon. The reason is that QIM embedding in YASS data embedding process introduces more zero coefficients.

As a summarization of this sub-section,we present a table (Table 1) to demonstrate the capacity and the outstanding feature of typical steganographic methods as well as their deviation in image statistics which are utilized by some targeted steganalytic methods.

4.2. **Universal Approaches.** Unlike specific steganalytic methods which require knowing the details of the targeted steganographic methods, universal steganalysis [66] requires less or even no such priori information. A universal steganalytic approach usually takes a learning based strategy which involves a training stage and a testing stage. The process is illustrated in Figure 12. During the process, a feature extraction step is used in both training and testing stage. Its function is to map an input image from a high-dimensional image space to a low-dimensional feature space. The aim of the training stage is to obtain a trained classifier. Many effective classifiers, such as Fisher linear discriminant (FLD), support vector machine (SVM), neural network (NN), etc., can be selected. Decision boundaries are formed by the classifier to separate the feature space into positive regions and negative regions with the help of the feature vectors extracted from the training images. In the testing stage, with the trained classifier that has the decision boundaries, an image under question is classified according to its feature vector's domination in the feature space. If the feature vector locates in a region where the classifier is labeled as positive, the testing image is classified as a positive class (stego image). Otherwise, it is classified as a negative class (cover image). Please note that some specific steganalytic methods may also take a similar learning based process. The difference between specific and universal methods lies in whether the features are effective in detecting a wide range of steganographic techniques. In the following, we mainly devote to presenting some typical universal steganalytic features.

4.2.1. *Image Quality Feature.* Steganographic schemes may more or less cause some forms of degradation to the image. Objective image quality measures (IQMs) are quantitative metrics based on image features for gauging the distortion. The statistical evidence left by

TABLE 1. Performance of typical steganographic methods

Steganography	Capacity*	Outstanding Features	Typical Deviated Statistics
LSB	1 bpp	substitute the least significant bit	"pairs of value" in histogram
BPCS	$\approx$ 4 bpp	substitute the noise-like binary image blocks	complexity histogram of the data-blocks
LSB Matching	1 bpp	plus or minus 1 randomly	histogram characteristic function
Stochastic Modulation	$\approx$ 0.8 bpp	modulate the embedded data as noise	pixel difference histogram
PVD	$>$ 1 bpp	embed data in the difference of neighboring pixel	"step effect" in pixel difference histogram
MBNS	$\approx$ 2 bpp	embed data in modulo value and the base value is determined adaptively	given any base value, more small symbols are generated
QIM/DM	depend on the specific application	quantizer is determined by message bit (usually in transform domain)	local correlation between coefficients
JSteg	$<$ 1 bpnc	substitute the least significant bit of JPEG DCT coefficients	"pairs of value" in DCT histogram
F5	$\approx$ 0.8 bpnc	decrease coefficients' absolute values and use matrix embedding	increased zero coefficients
OG	$\approx$ 0.4 bpnc	preserve the global DCT histogram	blockiness
MB	$\approx$ 0.8 bpnc	preserve the low-precision model	the high-precision bins follow the generalized Cauchy distribution too well
YASS	$<$ 0.4 bpnc	use randomized locations	more zero quantized coefficients are generated from P-region than I-region

* The capacity of some steganographic method may depend on the specific parameter and/or the specific image.

steganography may be captured by a group of IQMs and then exploited for detection[67]. In order to seek specific quality measures that are sensitive, consistent and monotonic to steganographic artifacts and distortions, the analysis of variance (ANOVA) technique is exploited and the ranking of the goodness of the metrics is done according to the F-score in the ANOVA tests. And the identified metrics can be defined as feature sets to distinguish between cover images and stego images.

4.2.2. *Calibration Based Feature.* Fridrich et al. [34] applied the feature-based classification together with the concept of calibration to devise a blind detector specific to JPEG images. Here the calibration means that some parameters of the cover image may be approximately recovered by using the stego image as side information. As a result, the

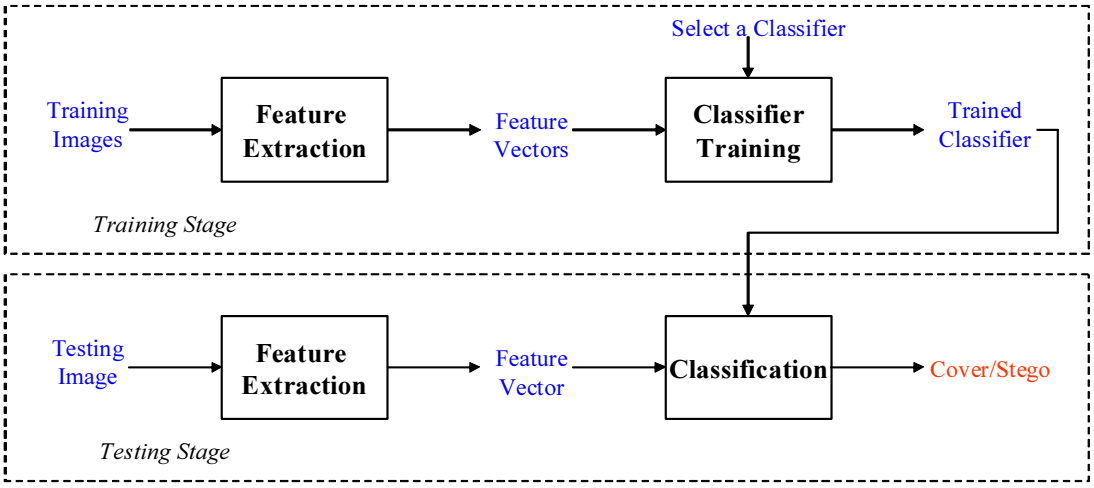


FIGURE 12. The process of a universal steganalytic method

calibration process increases the features' sensitivity to the embedding modifications while suppressing image-to-image variations.

In blind steganalysis scenarios, only the stego image J_1 can be obtained. By decompressing the stego image J_1 to the spatial domain, cropping by 4 pixels in each direction, and re-compressing with the same quantization table as J_1 , we may get a "calibrated" image J_2 with most macroscopic features similar to the original cover image. Instead of measuring the distance between the image and a statistical model, the distance between certain parameters of the image and the same parameters related to the recovered image are calculated and exploited for detection. 23 vector functionals $\mathbf{F}_i$ $(i = 1, 2, \cdots, 23)$ are applied to the stego JPEG image J_1. These 23 functionals include the global DCT coefficient histogram, individual histograms for 5 DCT modes $(h_{21}, h_{31}, h_{12}, h_{22}, h_{13})$, dual histograms for 11 DCT values $(-5, \cdots, 5)$, variation, L_1 and L_2 spital blockiness, and co-occurrence matrixes, etc. The same set of vector functionals $\mathbf{F}_i$ are then applied to J_2. The final feature f is obtained as an L_1 norm of the difference

$$f = \|\mathbf{F}_i(J_1) - \mathbf{F}_i(J_2)\|_{L_1} \tag{19}$$

where the L_1 norm is defined for a vector (or matrix) as a sum of absolute values of all vector (or matrix) elements.

By extending the 23 DCT feature set described previously, then applying calibration to the Markov process based features described in ref. [38] and reducing their dimension, Pevný et al. merged the resulting feature sets to produce a 274-dimensional feature vector [35]. The new feature set is then used to construct a multi-classifier capable of assigning stego images to six popular steganographic algorithms.

4.2.3. *Moment Based Feature.* The impact of steganography to a cover image can be regarded as introducing some stego-noise. As noise is added, some statistics of the image may be changed. It is effective to observe these changes in wavelet domain. Lyu and Farid [68] used the assumption that the PDF of the wavelet subband coefficients and that of the prediction error of the subband coefficients would change after data embedding. As a result, the statistical moments of the PDF (thereafter PDF moments), which can describe the PDF characteristics, were developed as steganalytic features. The n-th order PDF moment of a random variable S with a sequence of realizations $\{s_1, s_2, \cdots, s_N\}$ can

be computed as

$$M_n = E(S^n) = \frac{1}{N}\sum_{i=1}^{N}(s_i)^n \tag{20}$$

where $E(\cdot)$ is the expectation operator. In ref. [68], with a 3-level wavelet decomposition, the first four PDF moments, i.e., mean, variance, skewness, and kurtosis, of the subband coefficients at each high-pass orientation (horizontal, vertical and diagonal direction) of each level are taken into consideration as one set of features. The same kinds of PDF moments of the difference between the logarithm of the subband coefficients and the logarithm of the coefficients' cross-subband linear predictions at each high-pass orientation of each level are computed as another set of features. These two kinds of features provide satisfactory results when the embedding rate is high.

Goljan et al. [69] contributed a method with features from the first nine absolute central moments of the PDF (thereafter absolute PDF moments) of the estimated stego-noise. The n-th absolute PDF moment of a random variable S with the mean value $\overline{s}$ can be computed as

$$A_n = E(|S - \overline{s}|^n) = \frac{1}{N}\sum_{i=1}^{N}|s_i - \overline{s}|^n \tag{21}$$

And the estimation of stego-noise is performed in the wavelet domain with an adaptive denoising filter. It is expected that the features extracted from the estimated stego-noise will be more sensitive to data embedding and can greatly suppress the impact of the cover signal, compared to extracting features from cover signal directly. Besides, the stego-noise is only estimated in the one-level wavelet decomposition, justified by the fact that the SNR (stego-noise signal to cover image signal ratio) is high in this level. The features hit the right nail on the head and show superior performance in detecting additive steganography, even if the stego noise is weak.

As mentioned in Section 4.1.2, host-independent additive noise has a low-pass filtering effect on the PMF of the image [45]. The inverse Fourier transform of the PMF, also known as characteristic function (CF), will change accordingly. Xuan et al. [70] extended this conclusion to wavelet domain and used the statistical moments of the CF (thereafter CF moments) of the wavelet subband coefficients as steganalytic features. The n-th order CF moment is defined as

$$C_n = (\sum_{k=0}^{K/2}(k)^n|H(k)|)/(\sum_{k=0}^{K/2}|H(k)|). \tag{22}$$

where $H(k)$ is the discrete CF at frequency index k. The K-point discrete CF can be computed as

$$H(k) = \sum_{l=0}^{L-1} h(l)e^{\frac{j2\pi}{K}lk} \tag{23}$$

where $h(l)$ ($l \in \{0, ..., L-1\}$) is the normalized histogram of the coefficients, L is the total number of bins in the histogram, $K = 2^{\lceil \log_2 L \rceil}$, and $j = \sqrt{-1}$. The first three CF moments of the image and its three-level wavelet decomposited subband coefficients are used in ref. [70]. Improved from Xuan et al.'s work, Shi et al. [71] proposed to use a slightly different CF moment, which is defined as

$$C'_n = (\sum_{k=1}^{K/2}(k)^n|H(k)|)/(\sum_{k=1}^{K/2}|H(k)|). \tag{24}$$

The zero frequency component of the CF, i.e., $H(0)$ is deliberately excluded from eq. (22) for computing the new CF moment to enhance its discrimination capability. Besides, not only the CF moments of the the image and its wavelet subband coefficients are used, but also the CF moments of the prediction-error image, which is generated by a spatial prediction algorithm for removing the impact the image content, and its wavelet subband coefficients are also used as steganalytic features. This scheme is very sensitive to the changes caused by data hiding and outperforms the prior-arts.

Note that in practical computation, the PDF moments and absolute PDF moments can be directly calculated from the coefficient samples $\{s_1, s_2, \cdots, s_N\}$, as seen in eq. (20) and (21). But the CF moments, as in eq. (22) and (24), are computed from a histogram. Different histogram bin size may lead to different results if the data sample are in continuous values. In ref. [70, 71], Haar wavelet was used and therefore the coefficients are discrete. As a result, the bin of the histogram is easy to select for the discrete values. In general, the CF moment based features performs better than the PDF moment based features in most of the steganographic cases. The reason was first explained by Xuan et al. [72] and further verified by Wang et al. [73]. Simply speaking, when the energy of the stego-noise is low, the low-order PDF moments may not be able to catch the changes in PDF as effective as that low-order CF moments reflect the alterations in CF.

4.2.4. *Correlation Based Feature.* Data embedding may disturb the local correlation in an image. Here the correlation is mainly referred to the inter-pixel dependency for a spatial image, and the intra-block or inter-block DCT coefficient dependency for a JPEG image.

Sullivan et al. [74] modeled the inter-pixel dependency by Markov chain and depicted it by a gray-level co-occurrence matrix (GLCM) in practice. The element in the $(u+1)$-th row and the $(v+1)$-th column of the GLCM corresponds to the joint probability $P(X_i = u, X_{i-1} = v)$, where X_i denotes the i-th indexed pixel in an image $\mathbf{X}$, and $m, n \in \{0, 1, \cdots, 255\}$ for a 8-bit gray-scale image. For a cover image, the inter-pixel correlation is strong and thus the joint probability $P(X_i = u, X_{i-1} = v)$ is large. Therefore large values are mainly concentrated on the main diagonal of the GLCM and making it sparse. As the host-independent noise is added, large values in GLCM are spreading towards the minor diagonal direction in a stego image. Figure 13 illustrates the GLCM of a cover image, the GLCM of its stego image, and the difference between these two GLCMs. The joint probabilities on the main diagonal and near the main diagonal of the GLCM are served as steganalytic features. Although the features are not selected in a well-picked fashion from the GLCM, the method is in fact effective to a broad class of steganography, especially to the case of additive steganography. It shows that the i.i.d. assumption is unsuitable to characterize the cover data distribution for Alice, and Wendy can explore the data dependency for steganalysis.

Inspired by Sullivan et al.'s work, Shi et al. [38] proposed a Markov process based method that explores the intra-block DCT dependency for steganalyzing JPEG steganography blindly. In this method, a JPEG 2-D array is defined as the array consisting of the absolute values of the 8×8 block DCT coefficients that have been quantized by JPEG quantization steps but before a zig-zag scan and entropy encoding. Then four difference JPEG 2-D arrays are obtained by subtracting the JPEG 2-D array by its horizontal, vertical, main diagonal, and minor diagonal shift, respectively. Next, a threshold technique is taken to reduce the number of states (coefficient values) in the difference JPEG 2-D arrays. Specifically, the element in the array whose value is smaller than $-T$ or larger than T (e.g., $T = 5$) will be represented by $-T$ or T, respectively. Later, the transition probability matrix (TPM) is obtained for each difference JPEG 2-D array, and all transition probabilities are served as steganalytic features. The element in the $(T+1+u)$-th

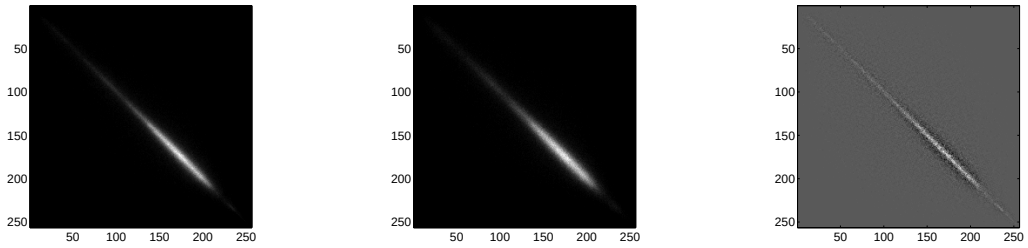


FIGURE 13. The GLCM of a cover image (a), the GLCM of a $\pm K$ ($K = 3$) stego image (b), and their difference (c).

row and the $(T+1+v)$-th column of the TPM corresponds to the conditional probability $P(F_i = u|F_{i-1} = v)$, where F_i denotes the i-th indexed coefficient in a difference JPEG 2-D array, and $u, v \in \{-T, -T+1, \cdots, 0, \cdots, T-1, T\}$. The difference JPEG 2-D array is expected to enlarge the data embedding disturbance and the threshold can reduce the feature dimension. In an updated work [39], in addition to the features with intra-block DCT dependency, the inter-block DCT dependency is also taken into account. A mode (or called sub-band) 2-D array is formed by re-arranging the JPEG 2-D array. And similar to the difference JPEG 2-D arrays, a horizontal mode 2-D array and a vertical difference mode 2-D array are generated and the threshold technique is exploited. By averaging the transition probability over 63 AC modes, the averaged probabilities in the TPM are served as inter-block features. The Markov process based features [38, 39] are very effective in detecting several JPEG steganographic scheme, even under a low embedding rate.

5. **The Continuing Competition.** From the development of the early LSB steganographic scheme[15] and Chi-square attatck[17], to the latest universal steganalytic methods[35] and YASS steganography[33] , it is not difficult to conclude that the fundamental relation between the research of steganography and steganalysis is their mutual resistance. The former tries to hide as large amount of information as possible while maintaining the undetectability level. And the later attempts to maximize the accuracy of detection in order to disable the steganography. Their competition is still going on.

5.1. **Improving Steganographic Security.** There are some factors that may influence the steganographic security, such as the number of changed pixels/coefficients, the amplitude of the stego-noise signal, the properties of cover images, etc. In the following we discuss some techniques for making the steganography less detectable.

5.1.1. *Increasing the Embedding Efficiency.* If cover images do not need to be modified at all for conveying secret information, certainly the warden cannot differentiate the cover images and stego images. Therefore, if the probability of modification to the images is less, the embedding changes to the image will reduce, and the security of the steganographic method may increase. Define the embedding efficiency as the number of embedded bits per one embedding change. Hence, increasing the embedding efficiency is a possible way to enhance the steganographic security. One technique, called matrix encoding [75, 30], can be used to increase the embedding efficiency. The concept was first proposed by Crandall [75] and implemented by Westfeld [30]. The basic idea is to divide coefficients into groups and use Hamming error correction codes to limit the changes in each group. A (d, n, k) code can be used to embed k bits into n coefficients by making at most d coefficients changed. The limitation of using Hamming code is that the embedding efficiency gets

high only when the embedding rate is low. Fridrich et al. [76] proposed to use random linear codes or simplex codes to cope with the case when the embedding rate is high. More advancement in constructing codes for improving embedding efficiency can be found in ref. [77, 78, 79].

5.1.2. *Reducing the Embedding Distortion.* Increasing the embedding efficiency can reduce the embedding changes to the image. However, it cannot guarantee that the distortion to the image is minimized. If not all of the coefficients are used for carrying data, Alice has the freedom to select the coefficients whose resultant distortions after data embedding are the smallest for modification. In this way, the stego image will be close to the cover image perceptually and statistically, thus enhancing the steganographic security. Perturbed quantization (PQ) steganography [34] is the first method addressing this issue. It is realized by changing some coefficients whose quantization errors are the smallest after data hiding. The method is facilitated by using the wet paper codes, a technique enabling Alice not to share the location of the changed coefficients with Bob. The method can be used in an information-reducing process which includes real transform and quantization, such as resizing and JPEG compression. Inspired by PQ steganography, Kim et al. [80] proposed the modified matrix encoding (MME) steganography by changing coefficients whose quantization errors plus embedding errors are the smallest when embedding data during the JPEG compression process. The method requires the uncompressed image as input and employs matrix encoding in embedding. Judging from the obtained results in ref. [81, 82], minimizing the embedding distortions does make the steganography less detectable. The tradeoff between embedding efficiency and embedding distortion is discussed in ref. [83].

5.1.3. *Selecting Proper Cover Images.* In some scenarios, Alice has the freedom to select the most unsuspicious stego images for conveying secret information. Kharrazi et al. [84] proposed a scheme for selecting the better images according to the availability of the knowledge of a potential steganalyzer. It implicitly assumes that the steganalyzer is not error free. If the steganalyzer is fully known, Alice can select the images which are undetectable by the steganalyzer. If Alice only has partially knowledge of the steganalyzer, for example, the input and output of the steganalyzer, she can choose the images which have similar properties with the undetectable images under some standard measures. If no knowledge of the steganalyzer is provided, Alice needs to decrease the possibility of being detectable by using the images with minimum changes.

5.2. **Enhancing Steganalytic Capability.** The statistics of stego images may be different from that of cover images. However, the deviated statistics may not obviously fall outside the normal scope where the statistics of cover images belong to. Therefore, some techniques may be needed to magnify the difference between cover and stego image (thereafter cover-and-stego difference) and thus enhancing the capability of a steganalyzer.

5.2.1. *Calibration – Estimating Cover Image's Statistics.* One way to magnify the cover-and-stego difference is to estimate the cover image's statistics from the testing image. The technique in using is often referred to as "calibration" and introduced in Section 4.2.2. By doing so, the estimated statistics can be employed to evaluate whether the statistics of the testing image are deviated. In a general case, denote the statistics in a vector form of the testing image as $\boldsymbol{F}_t$, and that of its cover image as $\boldsymbol{F}_c$. If the testing image is in fact a cover image, we will have $\|\boldsymbol{F}_t - \boldsymbol{F}_c\| = 0$, where $\|\cdot\|$ is the norm of a vector. Otherwise, we will have $\|\boldsymbol{F}_t - \boldsymbol{F}_c\| > 0$. But in practice Wendy does not know $\boldsymbol{F}_c$. She may estimate it from the testing image with some calibration techniques [62, 34, 46]. The estimated statistics vector is denoted as $\boldsymbol{F}_{\hat{c}(t)}$. It is expected that $\|\boldsymbol{F}_{t=c} - \boldsymbol{F}_{\hat{c}(t)}\| < \|\boldsymbol{F}_{t=s} - \boldsymbol{F}_{\hat{c}(t)}\|$, where

$\boldsymbol{F}_{t=c}$ and $\boldsymbol{F}_{t=s}$ denote the statistics of a testing image when it is a cover image and when it is a stego image, respectively. As discussed in Section 4.2.2, Fridrich [34] designed a powerful calibration method for JPEG images through cropping and re-compressing. Ker [46] proposed an effective scheme for estimating the statistics of spatial cover images via down-sampling. The calibration technique in essence provides a way to measure the scope the statistics of cover images belong to, therefore, it can enhance the capability of a steganalyzer. Obviously, the more accurate the estimation, the more accurate the steganalyzer with the calibration technique is.

5.2.2. *Re-embedding – Computing Re-stego Image's Statistics.* In contrast to the calibration technique which is independent of the steganographic method, another scheme that can magnify the cover-and-stego difference is related to the targeted steganographic algorithm. It is referred to as re-embedding and its operation is usually taken as embedding some arbitrary data into the testing image using the targeted steganographic algorithm. The resultant image is referred to as re-stego image. Re-embedding may be effective for only a limited number of steganographic methods which have a special property. That is, the re-embedding operation has a more severe impact on a cover image than on a stego image. For example, the LSB steganography with maximum embedding rate may make the unequalized value pair $2i$ and $2i+1$ in a cover image be balanced, while it cannot make the already equalized value pair in a stego image be more balanced. Denote the statistics of a re-stego image as $\boldsymbol{F}_{\hat{s}(t)}$. If the relation $\|\boldsymbol{F}_{t=c} - \boldsymbol{F}_{\hat{s}(t)}\| > \|\boldsymbol{F}_{t=s} - \boldsymbol{F}_{\hat{s}(t)}\|$ holds, the re-embedding takes effect and it can be used to enhance the capability of the steganalyzer. It has been successfully used in ref. [42, 50, 60, 61]. Note that calibration and re-embedding are not mutually opposed and they may work together to construct a steganalyzer, such as that used in attacking OutGuess [63].

5.2.3. *Filtering – Magnifying the Stego-noise.* Another way to magnify the cover-and-stego difference is to employ filtering, de-nosing, or prediction. The filtered/denoised/prediction residue may suppress the interference from the image content and magnify the stego-noise. Therefore, the statistics of the filtered residue from a stego image may be much different from the that from a cover image. It has been demonstrated in ref. [68, 71] that the methodology is very effective.

6. **Discussions and Conclusions.** In this paper, we review the fundamental concepts and notions as some typical techniques in steganography and steganalysis for digital images. Some developing trends of steganography are sketched as follows.

Adaptively selecting the embedding locations. We have witnessed plenty of steganographic methods [16, 85, 23, 24] using adaptive embedding strategy to embed data into the complex areas of an image, for the sake of avoiding causing perceptual artifacts. Besides, the edges and irregular texture areas may be hard to build a statistical model so that steganalytic method could be prone to make false decision. Therefore, selecting locations adaptively for embedding is still a promising solution in steganography. Note that the adaptive strategy should also be protected, such as using a key to ensure the randomness of the strategy. Or else the Wendy may use the same adaptive strategy to observe embedding artifacts.

Reducing embedding distortion and increasing embedding efficiency. It seems to be hard to preserve all statistics of the image after data embedding. Therefore, an intuitive idea is to minimize the embedding impact to the cover image, thus reducing the deviation of statistics. Through reducing the embedding changes and embedding energy, the stego image may be more similar to the cover image, both visually and statistically. Thus the statistics of the cover image may be preserved better.

Embedding data in the image creation process. If the data are embedded in an already generated image, it may be hard to preserve the image statistics. But what if data are embedded in the process of generating a new image? It has been shown that it's possible to improve the steganographic security by embedding data in the creation process of JPEG images [80, 34]. This may be a good solution for steganography.

Sacrificing the imperceptivity while preserving the statistics. Since a warden cannot obtain the original cover images and human eyes are insensitive to the distortion caused by steganography, it is feasible to introduce additional distortion in order to trade off a higher security level, such as YASS [33] .

Modern steganalytic techniques have greatly progressed. However, there are still some unsolved challenges. We summarize the future development directions of steganalysis in the following.

Identifying the type of steganography, the used parameters, the embedding rate, and even the embedding locations. If some more information can be dug out from the stego image, it will be more easy for Wendy to convict Alice. There are already some quantitative steganalytic techniques that can estimate the embedding rate [42, 86] and methods that can identify the steganography types [87, 88]. As long as the development of powerful steganalyzers, it may not be an illusion for Wendy to get the knowledge of the used parameters and even the embedding locations.

Working jointly with digital forensics. Some techniques used in steganalysis may be generalized in the applications of digital forensics [89], such as authenticating the origin, the generating process, or the doctoring evidences of digital media. In return, some digital forensics techniques may also help to stimulate the development of steganalysis.

From the ultimate competition between steganography and steganalysis, a byproduct, namely a natural image model, may be obtained, which is beneficial to both sides. For example, steganographic side can utilize the model to preserve image statistics, while steganalytic side can employ the model to examine if any statistic is deviated. It may also be useful in other related fields, such as digital forensics [89].

In summary, Alice wishes to safely send data to Bob as many as possible, while Wendy tries to neither malign an innocent cover medium nor let a single stego medium slip by. It seems that the competition between steganography and steganalysis will never end easily.

Acknowledgment. This work was supported by 973 program (2011CB302204),China and SZU R/D Fund (Project 201048). The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers.

REFERENCES

[1] R. J. Anderson and Fabien A. P. Petitcolas, On the limits of steganography, *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 4, pp. 474-481, 1998.

[2] Birgit P tzmann, Information hiding terminology-results of an informal plenary meeting and additional proposals, *Proc. of the First International Workshop on Information Hiding*, vol. 1174, pp. 347-350. Springer, 1996.

[3] Huaiqing Wang and Shuozhong Wang, Cyber warfare: Steganography vs. steganalysis, *Communications of the ACM*, vol. 47, no. 10, pp. 76-82, 2004.

[4] F. Neil Johnson and Sushil Jajodia, Steganalysis of images created using current steganography software. *Proc. of the Second International Workshop on Information Hiding*, vol. 1525 , pp. 273-273, 1998.

[5] Niels Provos and Peter Honeyman, Hide and seek: An introduction to steganography, *IEEE Security and Privacy*, vol. 1, no.3, pp. 32-44, 2003.

[6] R. Chandramouli, M. Kharrazi, and N. Memon, Image steganography and steganalysis concepts and practice, *Proc. of IWDW'03*, vol. 2939, pp. 35-49, Springer, 2003.

[7] Gustavus J. Simmons, The prisoners' problem and the subliminal channel, *Proc. of CRYPTO'83*, pp. 51-67, 1983.
[8] Tom Fawcett, Roc graphs: Notes and practical considerations for researchers, *Technical Report HPL-2003-4*, HP Laboratories, 2003.
[9] Christian Cachin, An information-theoretic model for steganography, Information and Computation, vol. 192, no. 1, pp. 41-56, 2004.
[10] Tomas Pevny and Jessica Fridrich, Benchmarking for steganography, *Proc. of the 10th International Workshop on Information Hiding*, vol. 5284, pp. 251-267, 2008.
[11] M. Wu, E. Tang, and B. Lin, Data hiding in digital binary image, *Proc. of 2000 IEEE International Conference on Multimedia and Expo*, vol. 1, pp. 393-396, 2000.
[12] G. Liang, S. Wang, and X. Zhang, Steganography in binary image by checking data-carrying eligibility of boundary pixels, *Journal of Shanghai University*, vol. 11, no. 3, pp. 272-277, 2007.
[13] F. Cayre and B. Macq, Data hiding on 3-d triangle meshes, *IEEE Trans. Signal Processing*, vol. 51, no. 4, pp. 939-949, 2003.
[14] Jessica Fridrich, Miroslav Goljan, and Rui Du, Reliable detection of lsb steganography in color and grayscale images. *Proc. of 2001 ACM workshop on Multimedia and security: new challenges*, pp. 27-30, ACM Press, 2001.
[15] W. Bender, D. Gruhl, N. Morimoto, and A. Lu, Techniques for data hiding, *IBM System Journal*, vol. 35, no. 3, pp. 313-336, 1996.
[16] Eiji Kawaguchi and Richard O. Eason, Principle and applications of bpcs steganography, *In Multimedia Systems and Applications*, vol. 3528, pp. 464-473, SPIE, 1998.
[17] A. Westfeld and A. Pfitzmann, Attacks on steganographic systems-breaking the steganographic utilities ezstego, jsteg, steganos, and s-tools-and some lessons learned. *Proc. of the 3rd Information Hiding Workshop*, vol. 1768, pp. 61-76, Springer, 1999.
[18] T. Sharp, An implementation of key-based digital signal steganography, *Proc. of the 4th Information Hiding Workshop*, vol. 2137, pp. 13-26, Springer, 2001.
[19] J. Mielikainen, Lsb matching revisited, *IEEE Signal Processing Letters*, vol. 13, no. 5, pp. 285-287, 2006.
[20] X. Li, B. Yang, D. Cheng, and T. Zeng, A generalization of lsb matching, *IEEE Signal Processing Letters*, vol. 16, no. 2, pp. 69-72, 2009.
[21] J. Fridrich, D. Soukal, and M. Goljan, Maximum likelihood estimation of secret message length embedded using pmk steganography in spatial domain, *Proc. of IST/SPIE Electronic Imaging: Security, Steganography, and Watermarking of Multimedia Contents VII*, vol. 5681, pp. 595-606, 2005.
[22] J. Fridrich and M. Goljan, Digital image steganography using stochastic modulation, *Proc. of IST/SPIE Electronic Imaging: Security and Watermarking of Multimedia Contents V*, vol. 5020, pp. 191-202, 2003.
[23] D. C. Wu and W. H. Tsai, A steganographic method for images by pixel-value diRerencing, *Pattern Recognition Letters*, vol. 24, no. 9-10, pp. 1613-1626, 2003.
[24] Xinpeng Zhang and Shuozhong Wang, Steganography using multiple-base notational system and human vision sensitivity, *IEEE Signal Processing Letters*, vol. 12, no. 1, pp. 67-70, 2005.
[25] B. Chen and G. W. Wornell, Quantization index modulation: A class of provably good methods for digital watermarking and information embedding, *IEEE Trans. Information Theory*, vol. 47, no. 4, pp. 1423-1443, 2001.
[26] H. Noda, M. Niimi, and E. Kawaguchi, High-performance jpeg steganography using quantization index modulation in dct domain, *Pattern Recognition Letters*, vol. 27, no. 5, pp. 455-461, 2006.
[27] J.J. Eggers, R. Bauml, R. Tzschoppe, and B. Girod, Scalar costa scheme for information embedding, *IEEE Trans. Signal Processing*, vol. 51, no. 4, pp. 1003-1019, 2003.
[28] Derek Upham, Jsteg, http://zooid.org/ paul/crypto/jsteg/.
[29] Allan Latham, Jphide, http://linux01.gwdg.de/ alatham/stego.html.
[30] Andrew Westfeld, F5-a steganographic algorithm: high capacity despite better steganalysis, *Proc. of the 4th Information Hiding Workshop*, vol. 2137, pp. 289-302, Springer, 2001.
[31] N. Provos, Defending against statistical steganalysis, *Proc. of the 10th USENIX Security Symposium*, pp. 323-325, 2001.
[32] P. Sallee, Model-based steganography, *Proc. of the 2nd Internation Workshop on Digital Watermarking*, vol. 2939, pp. 154-167, Springer, 2003.
[33] K. Solanki, A. Sarkar, and B. S. Manjunath, Yass: Yet another steganographic scheme that resists blind steganalysis,*Proc. of the 9th Information Hiding Workshop*, Springer, vol. 4567, pp. 16-31, 2007.

[34] J. Fridrich, Feature-based steganalysis for jpeg images and its implications for future design of steganographic schemes, *Proc. of the 6th Information Hiding Workshop*, Springer, vol. 3200 , pp. 67-81, 2004.

[35] Tomas Pevny and Jessica Fridrich, Merging markov and dct features for multi-class jpeg steganalysis. *Proc. of SPIE: Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents IX*, vol. 6505, pp. 3-14, 2007.

[36] Lifang Yu, Yao Zhao, Rongrong Ni, and Yun Q. Shi, A high-performance yass-like scheme using randomized big-blocks, *Proc. of the IEEE International Conference on Multimidea and Expo (ICME 2010)*, 2010.

[37] Fangjun Huang, Jiwu Huang, and Yun Qing Shi, An experimental study on the security performance of yass, *IEEE Trans. Information Forensics and Security*, vol. 5, no. 3, pp. 374-380, 2010.

[38] Y. Q. Shi, C. Chen, and W. Chen, A morkov process based approach to eRective attacking jpeg steganography. *Proc. of the 8th Information Hiding Workshop, Springer*, vol. 4437, pp. 249-264, 2006.

[39] C. Chen and Y. Q. Shi, Jpeg image steganalysis utilizing both intra-block and inter-block correlations, *Proc. of ISCAS'08*, pp. 3029-3032, 2008.

[40] Niels Provos and Peter Honeyman, Detecting steganographic content on the internet, *Proc. of NDSS'02: Network and Distributed System Security Symposium*, Internet Society, pp. 1-13, 2002.

[41] S. Dumitrescu, X. L. Wu, and Z. Wang, Detection of lsb steganography via sample pair analysis, *IEEE Trans. Signal Processing*, vol. 51, no. 7, pp. 1995-2007, 2003.

[42] J. Fridrich and M. Goljan, On estimation of secret message length in lsb steganography in spatial do- main. *Proc. of IST/SPIE Electronic Imaging: Security, Steganography, and Watermarking of Multimedia Contents VI*, vol. 5306, pp. 23-34, 2004.

[43] A. D. Ker, A general framework for the structural steganalysis of lsb replacement, *Proc. of the 7th Information Hiding Workshop*, Springer, vol. 3727, pp. 296-311, 2005.

[44] A. D. Ker, Fourth-order structural steganalysis and analysis of cover assumptions, *Proc. of IST/SPIE Electronic Imaging: Security, Steganography, and Watermarking of Multimedia Contents VIII*, vol. 6072, pp. 1-14, 2006.

[45] J. Harmsen and W. Pearlman, Steganalysis of additive noise modelable information hiding, *Proc. of IST/SPIE Electronic Imaging: Security, Steganography, and Watermarking of Multimedia Contents V*, vol. 5020, pp. 131-142, 2003.

[46] A. D. Ker. Steganalysis of lsb matching in grayscale images, *IEEE Signal Processing Letters*, vol. 12, no. 6, pp. 441-444, 2005.

[47] Xiaolong Li, Tieyong Zeng, and Bin Yang, Detecting lsb matching by applying calibration technique for diRerence image. *Proc. of the 10th ACM workshop on Multimedia and Security*, ACM Press, pp. 133- 138, 2008.

[48] J. Zhang, I. J. Cox, and G. Doerr, Steganalysis for lsb matching in images with high-frequency noise, *Proc. of IEEE Workshop on Multimedia Signal Processing*, pp. 385-388, 2007.

[49] G. Cancelli, G. Doerr, I. J. Cox, and M. Barni, Detection of +/-1 lsb steganography based on the amplitude of histogram local extrema, *Proc. of ICIP'08*, pp. 1288-1291, 2008.

[50] F. J. Huang, B. Li, and J. W. Huang, Attack lsb matching steganography by counting alteration rate of the number of neighbourhood gray levels, *Proc. of ICIP'07*, vol. 1, pp. 401-404, 2007.

[51] Jinggang Huang and David Mumford, Statistics of natural images and models. *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 541-547, 1999.

[52] J. H. He, J. W. Huang, and G. P. Qiu, A new approach to estimating hidden message length in stochastic modulation steganography, *Proc. of the 4th Internation Workshop on Digital Watermarking*, Springer, vol 3710, pp. 1-14, 2005.

[53] J. H. He and J. W. Huang, Steganalysis of stochastic modulation steganography, *Science in China Series: F-Information Sciences*, vol. 49, no. 3, pp. 273-285, 2006.

[54] M. Niimi, R. O. Eason, H. Noda, and E. Kawaguchi, Intensity histogram steganalysis in bpcs-steganography. *Proc. of IST/SPIE Electronic Imaging: Security and Watermarking of Multimedia Con- tents III*, vol. 4314, pp. 555-564, 2001.

[55] X. P. Zhang and S. Z. Wang, Vulnerability of pixel-value diRerencing steganography to histogram analysis and modication for enhanced security, *Pattern Recognition Letters*, vol. 25, no. 3, pp. 331-339, 2004.

[56] Bin Li, Yanmei Fang, and Jiwu Huang, Steganalysis of multiple-base notational system steganography, *IEEE Signal Processing Letters*, vol. 15, no. 493-496, 2008.

[57] K. Sullivan, Z. Bi, U. Madhow, S. Chandrasekaran, and B. S. Manjunath, Steganalysis of quantization index modulation data hiding, *Proc. of 2004 IEEE International Conference on Image Processing*, vol. 2, pp. 1165-1168, 2004.

[58] P. Guillon, T. Furon, and P. Duhamel, Applied public-key steganography, *IST/SPIE Electronic Imaging: Security and Watermarking of Multimedia Contents IV*, vol. 4675, pp. 38-49, 2002.

[59] Haz Malik, K. P. Subbalakshmi, and Rajarathnam Chandramouli, Steganalysis of qim-based data hiding using kernel density estimation. *Proc. of the 9th ACM Workshop on Multimedia and Security*, ACM Press, pp. 149-160, 2007.

[60] Haz Malik, Steganalysis of qim steganography using irregularity measure, *Proc. of the 10th ACM workshop on Multimedia and security*, ACM Press, pp. 149-158, 2008.

[61] H. Malik, K. P. Subbalakshmi, and R. Chandramouli, Non-parametric steganalysis of qim data hiding using approximate entropy, *Proc. of IST/SPIE Electronic Imaging: Security, Steganography, and Watermarking of Multimedia Content X*, vol. 6819, pp. 1-12, 2008.

[62] J. Fridrich, M. Goljan, and D. Hogea, Steganalysis of jpeg images: Breaking the f5 algorithm, *Proc. of the 5th Information Hiding Workshop*, Springer, vol. 2578, pp. 310-323, 2002.

[63] J. Fridrich, M. Goljan, and D. Hogea, Attacking the outguess, *Proc. of 2002 ACM Workshop on Multimedia and Security*, ACM Press, pp. 3-6, 2002.

[64] R. BAohme and A. Westfeld, Breaking cauchy model-based jpeg steganography with first order statistics, *Proc. of the 9th European Symposium On Research in Computer Security*, Springer, vol. 3193, pp. 125-140, 2004.

[65] Bin Li, Yun Q. Shi, and Jiwu Huang, Steganalysis of yass, *Proc. of the 10th ACM workshop on Multimedia and security*, ACM Press, pp. 139-148, 2008.

[66] X. Y. Luo, D. S. Wang, P. Wang, and F. L. Liu, A review on blind detection for image steganography, *Signal Processing*, vol. 88, no. 9, pp. 2138-2157, 2008.

[67] I. Avcibas, N. Memon, and B. Sankur, Steganalysis using image quality metrics, *IEEE Trans. Image Processing*, vol. 12, no. 2, pp. 221-229, 2003.

[68] Lyu Siwei and H. Farid, Detecting hidden message using higher-order statistics and support vector machines, *Proc. of the 5th Information Hiding Workshop*, Springer, vol. 2578, pp. 131-142, 2002.

[69] M. Goljan, J. Fridrich, and T. Holotyak, New blind steganalysis and its implications, *IST/SPIE Electronic Imaging: Security, Steganography, and Watermarking of Multimedia Contents VIII*, vol. 6072, pp. 1-13, 2006.

[70] G. R. Xuan, Y. Q. Shi, J. J. Gao, D. Zou, C. Y. Yang, Z. P. Zhang, P. Q. Chai, C. H. Chen, and W. Chen, Steganalysis based on multiple features formed by statistical moments of wavelet characteristic functions, *Proc. of the 7th Information Hiding Workshop*, Springer, vol. 3727, pp. 262-277, 2005.

[71] Y. Q. Shi, Guorong Xuan D. Zou, Jianjiong Gao, Chengyun Yang, Zhenping Zhang, Peiqi Chai, W. Chen, and C. Chen, Image steganalysis based on moments of characteristic functions using wavelet decomposition, prediction-error image, and neural network. *Proc. of IEEE Inter- national Conference on Multimedia and Expo*, IEEE Computer Society, vol. 1, pp. 269-272, 2005.

[72] Guorong Xuan, Jianjiong Gao, Y.Q. Shi, and D. Zou, Image steganalysis based on statistical moments of wavelet subband histograms in dft domain, *Proc. of IEEE International Workshop on Multimedia Signal Processing*, pp. 1-4, 2005.

[73] Y. Wang and P. Moulin, Optimized feature extraction for learning-based image steganalysis, *IEEE Trans. Information Forensics and Security*, vol. 2, no. 1, pp. 31-45, 2007.

[74] K. Sullivan, U. Madhow, S. Chandrasekaran, and B. S. Manjunath, Steganalysis for markov cover data with applications to images, *IEEE Trans. Information Forensics and Security*, vol. 1, no. 2, pp. 275-287, 2006.

[75] R. Crandall, Some notes on steganography, Posted on steganography mailing list, http://os.inf.tu-dresden.de/ westfeld/crandall.pdf, 1998.

[76] J. Fridrich and D. Soukal, Matrix embedding for large payloads, *IEEE Trans. Information Forensics and Security*, vol. 1, no. 3, pp. 390-395, 2006.

[77] W. M. Zhang, X. P. Zhang, and S. Z. Wang, A double layered plus-minus one data embedding scheme, *IEEE Signal Processing Letters*, vol. 14, no. 11, pp. 848-851, 2007.

[78] J. Fridrich, P. Lisonek, and D. Soukal, On steganographic embedding efficiency, *Proc. of the 8th Information Hiding Workshop*, Springer, no. 4437, pp. 282-296, 2007.

[79] JAurgen Bierbrauer and Jessica Fridrich, Constructing good covering codes for applications in steganography, *LNCS Trans. Data Hiding and Multimedia Security III*, vol. 4920, pp. 1-22, 2008.

[80] Y. Kim, Z. Duric, and D. Richards, Modified matrix encoding for minimal distortion steganography. *Proc. of the 8th Information Hiding Workshop, Springer*, vol. 4437, pp. 314-327, 2006.

[81] Jessica Fridrich, Tomas Pevny, and Jan Kodovsky, Statistically undetectable jpeg steganography: dead ends challenges, and opportunities, *Proc. of the 9th workshop on Multimedia and security* , ACM Press, pp. 3-14, 2007.

[82] Jan Kodovsky and J. Fridrich, Influence of embedding strategies on security of steganographic methods in the jpeg domain, *Proc. of IST/SPIE Electronic Imaging: Security, Forensics, Steganography, and Watermarking of Multimedia Contents X*, vol. 6819, pp. 1-13, 2008.

[83] Jessica Fridrich, Minimizing the embedding impact in steganography, *Proc. of the 8th ACM workshop on Multimedia and Security*, ACM Press, pp. 2-10, 2006.

[84] M. Kharrazi, H. T. Sencar, and N. Memon, Cover selection for steganographic embedding, *In Proc. of IEEE International Conference on Image Processing*, pp. 117-120, 2006.

[85] Grace Li, Nasir Memon, and R. Chandramouli, Adaptive steganography, *Proc. of SPIE: Se- curity and Watermarking of Multimedia Contents IV*, vol. 4675, pp. 69-78, 2002.

[86] Tao Zhang and Xijian Ping, A fast and eRective steganalytic technique against jsteg-like algorithms, *Proc. of ACM symposium on Applied computing*, ACM Press, pp. 307-311, 2003.

[87] Tomas Pevny and J. Fridrich, Multiclass blind steganalysis for jpeg images, *Proc. of IST/SPIE Electronic Imaging: Security, Steganography, and Watermarking of Multimedia Contents VIII*, vol. 6072, pp. 1-13, 2006.

[88] Tomas Pevny and J. Fridrich, Multiclass detector of current steganographic methods for jpeg format, *IEEE Trans. Information Forensics and Security*, vol. 3, no. 4, pp. 635-650, 2008.

[89] Weiqi Luo, Zhenhua Qu, Feng Pan, and Jiwu Huang, A survey of passive technology for digital image forensics, *Frontiers of Computer Science in China*, vol. 1, pp. 166-179, 2007.