Available online at www.sciencedirect.com

ScienceDirect

Signal Processing

Signal Processing 00 (2013) 1–13

Statistical Detection of Data Hidden in Least Significant Bits of Clipped Images ☆

Thanh Hai Thai*, Florent Retraint and Rémi Cogranne**

ICD - LM2S - Université de Technologie de Troyes - UMR STMR CNRS
12, rue Marie Curie - CS 42060 - 10004 Troyes cedex - France

Abstract

This paper studies the statistical detection of data hidden in the Least Significant Bits (LSB) plan of natural clipped images using the hypothesis testing theory. The main contributions are the following. First, this paper proposes to exploit the heteroscedastic noise model. This model, characterized by only two parameters, explicitly provides the noise variance as a function of pixel expectation. Using this model enhances the noise variance estimation and hence, allows the improving of detection performance of the ensuing test. Second, this paper introduces the clipped phenomenon caused by the limited dynamic range of the imaging device. Overexposed and underexposed pixels are statistically modeled and specifically taken into account to allow the inspecting of images with clipped pixels. While existing methods in the literature fail when the data is embedded in clipped images, the proposed detector still ensures a high detection performance. The statistical properties of the proposed GLRT are analytically established showing that this test is a Constant False Alarm Rate detector: it guarantees a prescribed false alarm probability.

Keywords: Hypothesis Testing, Natural Image Model, Optimal Detection, Steganalysis, Information Forensics.

1. Introduction

Steganography is the art and science of hiding communication. It aims to embed a secret message into host objects, or cover objects, to create so-called stego-objects. The stego-objects are then transmitted to the receiver via an insecure channel without raising suspicion of an adversary. The secret message can be retrieved by the receiver who knows in advance the hiding scheme and the secret key. In nowadays digital world, multimedia objects such as images, videos, and audios are often used as cover objects for steganography [5, 12, 22]. The communication channel provided by computer and network technologies helps to transmit the stego-objects easily and inconspicuously. Generally, steganographic techniques must satisfy two followings universal requirements [12] :

- The secret message must remain unchanged during or after the embedding process.
- The stego-object must remain unchanged or almost unchanged to the human eye.

The art of detecting hidden messages embedded in cover objects is called steganalysis. In the past decades, steganalysis has received a great attention from law enforcement agencies and media because the concept of steganography has been misused by anti-social elements and criminals over the internet [37]. Despite its intrinsic difficulty, the steganalysis is being developed and applied in many domains such as computer security, cyber warfare or national defence, to collect sufficient evidence about the presence of embedded message and to break the security of

*Corresponding author.

**With the financial support from the Prevention of and Fight against Crime Programme of the European Union European Commission - Directorate-General Home Affairs (2centre.eu project). Research partially funded by Troyes University of Technology (UTT) strategic program COLUMBO.

Email address: thanh_hai.thai@utt.fr (Thanh Hai Thai)

its carrier; see [37] and therein references for a survey of the methods available in the literature.

1.1. State of the Art

The paper concerns the image-based steganography and focuses on the popular technique of hiding information in the Least Significant Bit (LSB) of the cover image. Only the LSB replacement technique is studied. It is probably the oldest embedding technique in digital steganography. In fact, this method is simple, easy to implement and is used in about 70% of available steganographic software on the Internet [25]. In addition, the LSB replacement inspires the majority of other steganographic methods. Understanding the LSB replacement mechanism is a good starting point before addressing more complex data hiding schemes.

Steganalysis methods of LSB replacement can be divided into four categories:

1. Structural detectors exploit all combinatorial measures of the artificial dependence between sample differences and the parity structure of the LSB replacement in order to estimate the secret message length. Some representatives in this category are the Regular-Singular (RS) [24], the Sample Pair Analysis (SPA) [15], and the triple/quadruple [30].
2. Weighted Stego-Image (WS) detectors [23, 31] provide a computationally efficient estimate for the embedding rate of LSB replacement steganography which is formulated as a simple optimization problem. In contrast to structural detectors, the WS ones have left much space for improvement of designing a more accurate and reliable detector.
3. Statistical detectors [6, 9, 10, 13, 16, 34, 51] consider cover image pixels as realizations of random variables and exploit artifacts resulting from changes of their statistical properties in stego-images due to message embedding. These detectors provide a formal view on the information hiding problem by formulating it in the framework of hypothesis testing theory and analytically expressing the test performance.
4. Universal or blind detectors [25, 32, 38, 45, 49] employ machine learning methods based on a set of selected informative features to design an accurate classifier. Universal blind detectors are important because of their flexibility and ability to detect any steganographic scheme regardless of the embedding domain [37].

1.2. Limitations of Previous Approaches

In an operational context, for instance a steganalysis tool for law enforcement or intelligence agencies, the design of an accurate detector might not be sufficient. The most important and challenging problem is to provide a detector with analytically predictable results in order to guarantee a prescribed false alarm probability.

Both structural and WS detectors can provide overall acceptable detection performance. However, because these ad-hoc detectors have a limited exploitation of image model and statistical methods, their performance remains analytically unestablished. It is only evaluated on large databases.

Besides, as in all applications of machine learning, the main difficulties for blind detectors are the choice of appropriate feature set and the analytic establishment of detection performance. The latter remains an open problem in the framework of statistical learning [43]. In addition, it has been observed that blind detectors are subjected to the well-known cover source mismatch [2, 38] problem. Hence one needs images from the same camera to accurately train the classifier but such information might not be available or credible in practice.

On the opposite, by formulating the detection problem in the framework of hypothesis testing theory, only the statistical detectors proposed in [6, 9, 10, 16, 51] have been designed to warrant a prescribed false alarm probability. Unfortunately, these detectors lack an accurate model reflecting the statistical properties of natural images, even though adaptive regression models have been proposed in [9, 10, 16]. This leads to rather inaccurate estimation of both expectation and variance of pixels due to modeling errors. Besides, none of the prior detection schemes have introduced the clipping phenomenon which is due to the imaging system limited dynamic range. It has been observed [2] that using overexposed or underexposed areas of a natural image can significantly improve the detection performance but this has never been introduced in a detector.

1.3. Main Contributions of the Paper

The approach proposed in this paper involves the theory of statistical hypothesis testing [35]. It aims to design a statistical test for the steganalysis of LSB replacement scheme in a natural raw image. The main contributions are the following :

- Far from the Additive White Gaussian Noise (AWGN) model widely used in image processing, the proposed methodology is based on the heteroscedastic noise model which describes more accurately natural images [18, 20, 27]. The heteroscedastic property gives the variance of a pixel as a linear function of its expected value. This property, which has not yet studied in prior detectors, is considered in this paper to reduce estimation errors on expectation and variance of pixels.

- Instead of using a local regression model [6, 9, 10, 16, 51], this paper proposes to exploit a state-of-the-art denoising method [18]. This method is used together with the heteroscedastic noise model to significantly improve the estimation of pixels expectation and variance resulting in a higher detection performance.

- The clipping phenomenon, which makes the so-called *clipped* pixels being overexposed or underexposed, is taken into account in a rigorous statistical model. Based on this non-linear model of imaging device response, the paper provides a generalized accurate detector for any natural raw image (non-clipped or clipped). It is also shown that this phenomenon, which has never been taken into account in hidden data detection, may dramatically reduce the performance of prior-art detectors.

- Using an accurate statistical model of natural imag shown that the proposed test satisfies the Constan Alarm Rate (CFAR) property. This test allows the anteeing of a prescribed false-alarm probability a setting of decision threshold independently of the content while other statistical detectors [16] fail i tice.

It should also be highlighted that the proposed metho relies on general statistical concepts and general prope natural images. It can thus be applied to more comple hiding schemes: a first step for the statistical detection c matching scheme has been proposed in [7, 8].

1.4. Organization of the Paper

The paper is organized as follows. Section 2 states the information hiding detection problem in the framework of statistical hypothesis testing theory based on the heteroscedastic noise model. Section 3 proposes a Generalized Likelihood Ratio Test (GLRT) when the cover image content parameters are unknown. Maximum Likelihood (ML) estimates of parameters are provided and replaced in the Likelihood Ratio function. Section 4 addresses a GLRT taking into account the impact of clipping phenomenon. This detector serves as an upper bound of the previously proposed detector or any practical detector which does not regard this kind of degradation. Section 5 studies numerical performances of the proposed detector on synthetic and real natural images and presents some comparisons with other state-of-the-art detectors. Finally, Section 6 concludes the paper.

2. Problem Statement

This section describes the heteroscedastic noise model by relying on the image acquisition pipeline. It also formulates the problem of steganalysis in the framework of hypothesis testing theory.

2.1. Heteroscedastic Noise Model

This paper deals with natural images which are acquired by a digital imaging sensor. Let us assume that a natural raw image before quantization is a vector $\mathbf{Y} = \{y_i\}_{i\in\mathcal{I}}$ of N pixels where $\mathcal{I} = \{1, \dots, N\}$ is the index set of pixels. The raw image can be modeled by considering the noises that contribute to the degradation of the captured image during the image acquisition process [46, 48, 27, 20]. In fact, the photon shot noise and dark current are modeled as Poissonian random variables whereas other electronic noises are modeled as a zero-mean Gaussian one. For the sake of simplification, the normal approximation of the Poisson distribution can be exploited due to the large number of counted electrons. Finally, the raw pixel y_i follows the Gaussian distribution [46, 48]:

$$y_i \sim \mathcal{N}(\mu_i, a\mu_i + b) \tag{1}$$

where μ_i denotes the expectation of raw pixel y_i and (a, b) characterize a camera model. The raw pixel y_i follows a Gaussian distribution whose probability density function (pdf) $f_{\theta_i}(x)$ is entirely characterized by the parameter vector $\boldsymbol{\theta}_i = (\mu_i, a, b)^T$ where $\mathbf{U}^T$ denotes the transpose of the matrix $\mathbf{U}$.

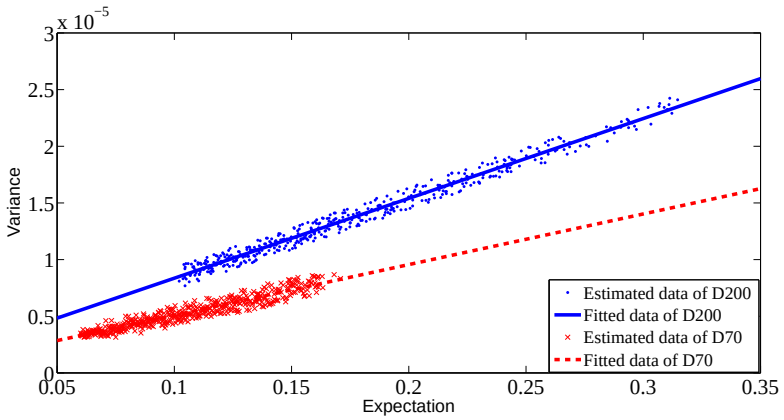

Figure 1: Heteroscedastic relation between expectation and variance of a natural image.

The model (1) shows that the noise variance is characterized as a linear function of the pixel's mathematical expectation. An example is given in Figure 1 to illustrate both the heteroscedastic relation and the camera model dependence of parameters a and b for two *Nikon* reflex cameras. In fact, the AWGN is not appropriate to model accurately a natural image due to the dominant contribution of the Poisson noise corrupting the raw pixel. It is proposed in this paper to use the heteroscedastic noise model (1) to achieve better estimation of pixels' expectation and variance, instead of the AWGN model. Moreover, it is important to remind that this paper focuses on addressing the steganalysis of LSB replacement scheme in any natural image, in particular clipped image. The heteroscedastic noise model is more appropriate to deal with clipped images than the AWGN. Besides, it should be noted that the parameters (a, b) mainly depend on ISO sensitivity [20, 47]. However, this dependance does not affect the detection performance of the proposed detector, see more results in Section 5.2.

2.2. Statistical Model for LSB Replacement

The LSB replacement mechanism was statistically described in [13]. First, let define a vector $\mathbf{C} = \{c_i\}_{i\in\mathcal{I}}$ representing a cover image coded with B bits, i.e., $c_i \in \{0, 1, \dots, l\}$ where $l = 2^B - 1$, in the following manner $c_i = \lfloor y_i \rfloor$ where $\lfloor\cdot\rfloor$ is the operation of uniform quantization with a unitary step. According to [6, 13], the probability distribution P_{θ_i} of the cover pixel c_i can be simplified as

$$P_{\theta_i} = \Big\{p_{\theta_i}[0], p_{\theta_i}[1], \dots, p_{\theta_i}[l]\Big\} \tag{2}$$

where $\forall k \in \{0, \dots, l\}$

$$\begin{aligned} p_{\theta_i}[k] &= \frac{1}{\sigma_i}\int_{k-\frac{1}{2}}^{k+\frac{1}{2}} \phi\Big(\frac{x-\mu_i}{\sigma_i}\Big)\,dx \\ &= \frac{1}{\sigma_i}\phi\Big(\frac{k-\mu_i}{\sigma_i}\Big) + o(\sigma_i^{-2}) \end{aligned} \tag{3}$$

Here, the variance is given by $\sigma_i^2 = a\mu_i + b$, see Equation (1), $\phi(x)$ and $\Phi(x)$ denote respectively the pdf and the cumulative distribution function (cdf) of the standard Gaussian distribution. It is assumed in the first stage that the saturation is absent, i.e. the probability that one observation y_i exceeds over the boundary 0 or l is negligible. The term $o(\sigma_i^{-2})$ that represents an error of approximation is negligible provided that the quantization step is very small in comparison with the noise in natural raw images. This consideration is quite realist because a natural raw image is often coded with $B \in \{12, 14, 16\}$ bits.

The LSB replacement technique is performed by replacing a ratio R of the cover pixels LSBs by the encrypted secret message to create a stego-image $\mathbf{S} = \{s_i\}_{i\in\mathcal{I}}$. The insertion is assumed to be equiprobable for every cover pixel and the hidden bits are independent and identically distributed [13, 23]. From the total probability theorem, the probability distribution of the stego-pixel s_i at rate R, denoted $Q_{R,\boldsymbol{\theta}_i}$, is given by [13]

$$Q_{R,\boldsymbol{\theta}_i} = \Big\{q_{R,\boldsymbol{\theta}_i}[0], q_{R,\boldsymbol{\theta}_i}[1], \ldots, q_{R,\boldsymbol{\theta}_i}[l]\Big\} \tag{4}$$

where

$$q_{R,\boldsymbol{\theta}_i}[k] = \frac{R}{2}\Big(p_{\boldsymbol{\theta}_i}[k] + p_{\boldsymbol{\theta}_i}[\bar{k}]\Big) + (1-R)p_{\boldsymbol{\theta}_i}[k], \tag{5}$$

where $\bar{k}$ indicates the integer k with LSB flipped $\bar{k} = k + (-1)^k$.

2.3. Hypothesis Testing Formulation

The hidden bits detection problem is cast in the framework of hypothesis testing theory [35]. It relies on the heteroscedastic noise model (1) of natural raw images. When inspecting the image $\mathbf{Z} = \{z_i\}_{i\in\mathcal{I}}$ that is either a cover image or a stego-one, the goal of the test is to decide between two hypotheses defined by

$$\begin{cases} \mathcal{H}_0 &= \Big\{z_i \sim P_{\boldsymbol{\theta}_i}, \forall i \in \mathcal{I}, \boldsymbol{\theta}_i \in \mathbb{R}^3\Big\} \\ \mathcal{H}_1 &= \Big\{z_i \sim Q_{R,\boldsymbol{\theta}_i}, \forall i \in \mathcal{I}, \boldsymbol{\theta}_i \in \mathbb{R}^3, R \in (0,1]\Big\} \end{cases} \tag{6}$$

As previously explained, this paper focuses on guaranteeing a prescribed false-alarm probability. Hence, let

$$\mathcal{K}_{\alpha_0} = \Big\{\delta : \sup_{\boldsymbol{\theta}} \mathbb{P}_{\mathcal{H}_0}\Big[\delta(\mathbf{Z}) = \mathcal{H}_1\Big] \leq \alpha_0\Big\}$$

be the class of tests with a false alarm probability upper-bounded by α_0. Here $\mathbb{P}_{\mathcal{H}_0}(E)$ stands for the probability of event E under hypothesis $\mathcal{H}_0$, and the supremum over $\boldsymbol{\theta}$ has to be understood as whatever model parameters might be. Among all the tests in $\mathcal{K}_{\alpha_0}$, it is aimed at finding a test δ which maximizes the power function, defined by the correct detection probability:

$$\beta_\delta = \mathbb{P}_{\mathcal{H}_1,R}\Big[\delta(\mathbf{Z}) = \mathcal{H}_1\Big], \tag{7}$$

where $\mathbb{P}_{\mathcal{H}_1,R}(E)$ stands for the probability of event E under hypothesis $\mathcal{H}_1$ for a given embedding rate R.

The problem (6) involves two main difficulties. First, the hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ are composite because the embedding rate R is unknown. Second, the nuisance parameters $\boldsymbol{\theta}_i$ are unknown in practice.

For the first difficulty, there is no algorithm in the literature that can establish a rigorous framework of estimating the embedding rate R, at least by ML approach. Even when all the parameters $\boldsymbol{\theta}_i = (\mu_i, a, b)^T$ are known, the problem with only one unknown parameter R remains practically unsolvable. The detector proposed in [51] attempted to reduce the problem with two composite hypotheses to the one with two simple hypotheses based on a local asymptotic approach [33]. This detector however has shown a considerable loss of optimality.

This paper assumes that the embedding rate R is known in advance and addresses the second difficulty. A possible approach to deal with unknown nuisance parameters consists in eliminating them by using the invariance principle [35]. This approach has been discussed in [21, 42] and has achieved a good performance in some applications [17]. However, in the heteroscedastic noise model, the image parameter μ_i appears in both mathematical expectation and variance of the pixel z_i. The invariant approach may not be applied due to a difficulty of finding a group of transformation under which the problem remains invariant. Another approach is to design a GLRT by replacing the unknown parameters by ML estimates [39, 4]. The main challenge of this approach is to provide accurate ML estimates of nuisance parameters with statistical properties in order to analytically establish the detection performance of the GLRT.

In a practical context, the camera parameters (a, b) should be estimated from a single image. The ML estimation of camera parameters (a, b) was first proposed in [20]. However, this method relies on solving numerically the maximization of likelihood function by using the Nelder-Mead optimization method [36], which leads to an impossibility of analytically establishing the statistical properties of ML estimates. Another approach is based on Weighted Least Square (WLS) estimation [47], which are asymptotically equivalent to the ML estimates in large samples. This approach allows to provide statistical properties of WLS estimates, resulting in establishing analytically the detection performance of a GLRT. This paper mainly focuses on dealing with unknown image parameters μ_i assuming that the camera parameters (a, b) and the rate R are known in order to designing a GLRT for the steganalysis of LSB replacement in any natural raw image (non-clipped or clipped images). The design of a GLRT taking into account the variance of WLS estimates is beyond the scope of the paper.

3. Generalized Likelihood Ratio Test with Unknown Image Parameters

This section addresses a GLRT in which the image parameters μ_i are unknown. It is assumed that the camera parameters (a, b) and the hiding rate R are known in advance. In this section, the paper assumes that the saturation is absent, i.e. there is no clipping in the inspected image. The GLRT for clipped images will be studied in Section 4.

3.1. Design of the Proposed GLRT

It can be noted that the number of unknown nuisance parameters μ_i grows with the number of pixels. To reduce the

difficulty of unknown parameter estimation, the approach proposed in this paper consists in segmenting the inspected image $\mathbf{Z}$ into K non-overlapping segments, denoted S_k, of size n_k, $k \in \{1, \dots, K\}$. Each segment S_k is represented by a vector $\mathbf{z}_k = \{z_{k,i}\}_{i=1}^{n_k}$ where pixels share the same mean μ_k. Let $\boldsymbol{\theta}_k = (\mu_k, a, b)^T$ be the parameter vector characterizing the segment S_k. Therefore, the unknown image parameters μ_k can be estimated by the ML estimation approach in each segment.

The Likelihood Ratio (LR) $\Lambda(z_{k,i})$ of one observation $z_{k,i}$ is given by

$$\begin{aligned}\Lambda(z_{k,i}) &= \log \frac{q_{R,\boldsymbol{\theta}_k}[z_{k,i}]}{p_{\boldsymbol{\theta}_k}[z_{k,i}]} \\ &= \log\Big[1 + R\Big(\exp\Big(\frac{\gamma_{k,i}(z_{k,i} - \mu_k)}{2\sigma_k^2}\Big) - 1\Big) + o(\sigma_k^{-2})\Big] \end{aligned} \tag{8}$$

where $\gamma_{k,i} = z_{k,i} - \bar{z}_{k,i}$, $\sigma_k^2 = a\mu_k + b$. The random variable $\gamma_{k,i}$ represents the impact of insertion and takes value 1 and -1 depending on the parity of $z_{k,i}$. By using the first-order series Taylor expansion of $\log(1 + x)$ and $\exp(x)$, one obtains

$$\Lambda(z_{k,i}) = \frac{R}{2\sigma_k^2}\gamma_{k,i}(z_{k,i} - \mu_k) + o(\sigma_k^{-2}). \tag{9}$$

In practice the expectation μ_k and the variance σ_k^2 of each segment S_k are unknown. In such situation, a usual solution consists in replacing the unknown parameters by their ML estimates in (9). This leads to the GLRT defined by

$$\widehat{\delta} = \begin{cases} \mathcal{H}_0 \quad \text{if} \quad \widehat{\Lambda}(\mathbf{Z}) = \displaystyle\sum_{k=1}^{K}\sum_{i=1}^{n_k} \widehat{\Lambda}(z_{k,i}) < \widehat{\tau} \\ \mathcal{H}_1 \quad \text{if} \quad \widehat{\Lambda}(\mathbf{Z}) = \displaystyle\sum_{k=1}^{K}\sum_{i=1}^{n_k} \widehat{\Lambda}(z_{k,i}) \geq \widehat{\tau} \end{cases} \tag{10}$$

where, to ensure the GLRT to be in the class $\mathcal{K}_{\alpha_0}$, the decision threshold $\widehat{\tau}$ is the solution of the equation

$$\mathbb{P}_{\mathcal{H}_0}\Big[\widehat{\Lambda}(\mathbf{Z}) \geq \widehat{\tau}\Big] = \alpha_0 \tag{11}$$

and the Generalized Likelihood Ratio (GLR) $\widehat{\Lambda}(z_{k,i})$ of one observation $z_{k,i}$ is defined by

$$\widehat{\Lambda}(z_{k,i}) = \frac{1}{\hat{\sigma}_k^2}\gamma_{k,i}(z_{k,i} - \hat{\mu}_k) \tag{12}$$

3.2. Segmentation and Estimation of Cover Image Parameters

The segmentation used in this paper is based on the denoising of natural raw images. The idea is that pixels whose denoised value takes the same grayscale level are independent and identically distributed. This is due to the fact that noise variance depends on pixel's expectation. Hence, pixels with the same expectation also share the same variance. The heteroscedastic noise model needs to be taken into account in the denoising algorithm. This paper designs a denoising filter applied to non-clipped images by combining a homomorphic transformation and a filter for homoscedastic noise. This filter is detailed in Appendix B.

Formally, let $\mathfrak{D}$ be the ideal denoising filter. It means that the image content of a noisy non-clipped image can be very accurately recovered

$$\mathfrak{D}(z_i) = \mathbb{E}[z_i] = \mu_i. \tag{13}$$

Each segment S_k, which is characterized by its central value u_k and allowed deviation $\Delta_k > 0$, is defined as :

$$S_k = \Big\{z_i : \mathfrak{D}(z_i) \in \Big[u_k - \frac{\Delta_k}{2}, u_k + \frac{\Delta_k}{2}\Big[, i \in \mathcal{I}\Big\}. \tag{14}$$

In other words, the dynamic range of the image is uniformly divided into K intervals of length Δ_k. The number of segment K used in this paper is set to the number of quantization levels, e.g. $K = 2^B$ and $\Delta_k = 1$. Consequently, the ML estimate of the image parameter μ_k in the segment S_k is given by

$$\hat{\mu}_k = \frac{1}{n_k}\sum_{i=1}^{n_k} z_{k,i}. \tag{15}$$

This ML estimate is unbiased and follows the Normal distribution

$$\hat{\mu}_k \sim \mathcal{N}\Big(\mu_k, \frac{\sigma_k^2}{n_k}\Big) \tag{16}$$

where $\sigma_k^2 = a\mu_k + b$.

$$\hat{\sigma}_k^2 \sim \mathcal{N}\Big(\sigma_k^2, \frac{a^2}{n_k}\sigma_k^2\Big). \tag{17}$$

3.3. Statistical Performance of the GLRT $\widehat{\delta}$

In order to establish the statistical performance of the test, it is necessary to characterize the distribution of the GLR $\widehat{\Lambda}(\mathbf{Z})$ under each hypothesis $\mathcal{H}_j$, $j \in \{0, 1\}$. Because of a large number of pixels N in any natural raw image, its statistical distribution is given in virtue of the Lindeberg Central Limit Theorem (CLT) [35, theorem 11.2.5] (see details in Appendix A)

$$\widehat{\Lambda}(\mathbf{Z}) \xrightarrow{D} \mathcal{N}(m_j, v_j) \tag{18}$$

where $\xrightarrow{D}$ represents the convergence in distribution and (m_j, v_j) denote respectively the mean and variance of the GLR $\widehat{\Lambda}(\mathbf{Z})$ under each hypothesis $\mathcal{H}_j$

$$m_0 = 0 \tag{19}$$

$$v_0 = \sum_{k=1}^{K}\frac{n_k + 1}{\sigma_k^2 + \frac{a^2}{n_k}} \tag{20}$$

$$m_1 = \frac{R}{2}\sum_{k=1}^{K}\frac{n_k}{\sigma_k^2} = \frac{R}{2}\sum_{k=1}^{K}\frac{n_k}{a\mu_k + b} \tag{21}$$

$$v_1 = \sum_{k=1}^{K}\Big[\frac{n_k + 1 + \frac{n_k R}{2\sigma_k^2}}{\sigma_k^2 + \frac{a^2}{n_k}} - \frac{R^2 n_k}{4\sigma_k^4}\Big]. \tag{22}$$

The parameters (a, b) characterizing camera models also appear in the formula of (m_j, v_j). Because a natural image is heterogenous, it is crucial to normalize the GLR $\widehat{\Lambda}^\star(\mathbf{Z})$ in order to set the

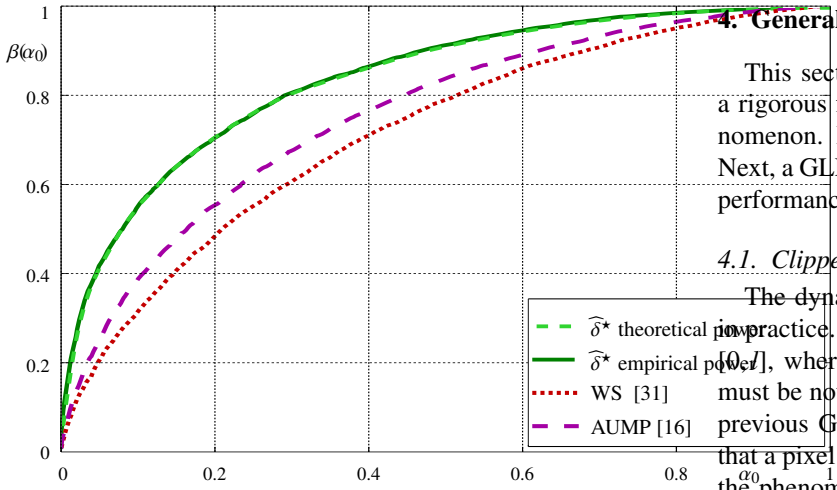

Figure 2: Detection performance on a non-clipped image: simulated data.

threshold independently of the image content. The normalized GLR $\widehat{\Lambda}^\star(\mathbf{Z})$ is defined by

$$\widehat{\Lambda}^\star(\mathbf{Z}) = \frac{\widehat{\Lambda}(\mathbf{Z})}{\sqrt{v_0}} = \sum_{k=1}^{K}\sum_{i=1}^{n_k} \frac{1}{\hat{\sigma}_k^2\sqrt{v_0}}(z_{k,i} - \bar{z}_{k,i})(z_{k,i} - \hat{\mu}_k). \quad (23)$$

However, the variance v_0 is not defined since the image parameters μ_k are unknown in practice. It is proposed to replace μ_k by $\hat{\mu}_k$ in (20) to obtain an estimate of v_0, denoted $\hat{v}_0$. Accordingly, the corresponding test $\widehat{\delta}^\star$ is rewritten as follows

$$\widehat{\delta}^\star = \begin{cases} \mathcal{H}_0 & \text{if} \quad \widehat{\Lambda}^\star(\mathbf{Z}) < \widehat{\tau}^\star \\ \mathcal{H}_1 & \text{if} \quad \widehat{\Lambda}^\star(\mathbf{Z}) \geq \widehat{\tau}^\star \end{cases} \quad (24)$$

From the Slutsky's theorem [35, Theorem 11.2.11], one obtains straightforwardly the decision threshold and the power function of the test $\widehat{\delta}^\star$

Theorem 1. *For any non-clipped natural raw image whose parameters are unknown, the decision threshold of the test $\widehat{\delta}^\star$ based on the decision function $\widehat{\Lambda}^\star(\mathbf{Z})$ is given by :*

$$\widehat{\tau}^\star = \Phi^{-1}(1-\alpha_0). \quad (25)$$

Theorem 2. *The power function of the test $\widehat{\delta}^\star$ is given by :*

$$\beta_{\widehat{\delta}^\star} = 1 - \Phi\left(\frac{\widehat{\tau}^\star\sqrt{v_0} - m_1}{\sqrt{v_1}}\right). \quad (26)$$

Similar to the AUMP test[16], the test $\widehat{\delta}^\star$ satisfies the CFAR property because of the theorem 1. However, the AUMP test showed a loss of optimality, even in a simulated data, due to the use of local regression model and AWGN model. The main advantage of the proposed test $\widehat{\delta}^\star$ is that there is no bias of the estimate $\widehat{\Lambda}^\star(\mathbf{Z})$, which results in no loss of optimality. Its power function $\beta_{\widehat{\delta}^\star}$ highly depends on the hiding rate R and the camera parameters (a, b). The smaller the rate R is or the more important the noise variance characterized by parameters (a, b) is, the smaller the detectability is: a change in the LSB is not significant compared to the noise level.

4. Generalized Likelihood Ratio Test in Clipped Images

This section extends the previous approach by establishing a rigorous framework to resolve the problem of clipping phenomenon. First, the clipped image model is briefly presented. Next, a GLRT is designed based on this model and its statistical performance is analytically provided.

4.1. Clipped Image Model

The dynamic range of acquisition system is always limited in practice. Indeed, recorded pixels are bounded by the interval $[0, l]$, where $l = 2^B - 1$, if an image is coded with B bits. It must be noted that the heteroscedastic noise model used in the previous GLRT works on the assumption that the probability that a pixel y exceeds over the boundary 0 or l is negligible, i.e. the phenomenon of clipping does not happen in the inspected image. In practice, this assumption is not always true, particularly if noise is very important, which leads to cause the pixel y to exceed these bounds. In the digital imaging system, the recorded pixels exceeding theses bounds are replaced by the bounds themselves. Accordingly, the *clipped* pixel $\tilde{y}_i$ is defined by

$$\tilde{y}_i = \max(0, \min(y_i, l)), \quad \forall i \in \mathcal{I} \quad (27)$$

where y_i follows the heteroscedastic noise model (1). As explained in [18], the generalized pdf of clipped pixel $\tilde{y}_i$ is defined by

$$\tilde{f}_{\theta_i}(x) = \Phi\Big(\frac{-\mu_i}{\sigma_i}\Big)\delta_0(x) + \frac{\frac{1}{\sigma_i}\phi\Big(\frac{x-\mu_i}{\sigma_i}\Big)}{\Phi\Big(\frac{l-\mu_i}{\sigma_i}\Big) - \Phi\Big(\frac{-\mu_i}{\sigma_i}\Big)} + \Phi\Big(\frac{\mu_i - l}{\sigma_i}\Big)\delta_0(l-x). \quad (28)$$

where δ_0 is the Dirac delta impulse at 0. The first and the last term in (28) correspond to the probabilities of clipping from below and from above while the second term represents the pdf of a truncated Gaussian random variables (see details about the truncated Gaussian distribution in [29]). The clipping phenomenon presents two main difficulties. On the one hand, it modifies the statistical distribution of raw pixels, which now follow the double-censored Gaussian distribution. On the other hand, the variance can no longer be treated as the linear function of the mathematical expectation. The GLRT proposed in the previous section 3 is no longer relevant when dealing with clipped images. The estimation methodology of image parameters μ_i also needs to be properly modified.

4.2. Design of the Proposed GLRT

Now the cover image and stego image are respectively denoted by $\tilde{\mathbf{C}} = \{\tilde{c}_i\}_{i\in\mathcal{I}}$ and $\tilde{\mathbf{S}} = \{\tilde{s}_i\}_{i\in\mathcal{I}}$. Here again, without taking into account the impact of quantization, the probability distribution $\tilde{P}_{\theta_i} = \left\{\tilde{p}_{\theta_i}[0], \tilde{p}_{\theta_i}[1], \ldots, \tilde{p}_{\theta_i}[l]\right\}$ of the cover pixel $\tilde{c}_i$ can be defined by (28). Due to the impact of hidden data (5), the probability distribution $\tilde{Q}_{R,\theta_i}$ of the stego-pixel $\tilde{s}_i$ is modified as $\tilde{Q}_{R,\theta_i} = \left\{\tilde{q}_{R,\theta_i}[0], \tilde{q}_{R,\theta_i}[1], \ldots, \tilde{q}_{R,\theta_i}[l]\right\}$ where

$$\tilde{q}_{R,\theta_i}[k] = \frac{R}{2}\Big(\tilde{p}_{\theta_i}[k] + \tilde{p}_{\theta_i}[\bar{k}]\Big) + (1-R)\tilde{p}_{\theta_i}[k]. \quad (29)$$

Let $\tilde{\mathbf{Z}} = \{\tilde{z}_i\}_{i\in\mathcal{I}}$ be the inspected image. The image $\tilde{\mathbf{Z}}$ is segmented into K non-overlapping segments S_k of size n_k, $k \in \{1,\dots,K\}$, defined by

$$S_k = \Big\{\tilde{z}_i : \tilde{\mathfrak{D}}(\tilde{z}_i) \in \Big[u_k - \frac{\Delta_k}{2}, u_k + \frac{\Delta_k}{2}\Big[, i \in \mathcal{I}\Big\}. \tag{30}$$

where $\tilde{\mathfrak{D}}$ be the ideal denoising and declipped operator

$$\tilde{\mathfrak{D}}(\tilde{z}) = \mu. \tag{31}$$

It must be noted that when dealing with a clipped image, it is of crucial importance to understand the relation between the clipped data and non-clipped data. The output image after the denoising filter can only be treated as an estimate of $\mathbb{E}[\tilde{z}]$, not μ. It is proposed in [18] to perform a declipped operator, i.e. an inverse transformation to obtain an estimate of μ. As demonstrated in Appendix C, the ML estimate $\hat{\mu}_k$ follows the asymptotic Gaussian distribution

$$\hat{\mu}_k \sim \mathcal{N}\Big(\mu_k, \frac{1}{n_k\mathcal{F}(\mu_k)}\Big) \tag{32}$$

where $\mathcal{F}$ is the Fisher information defined in Equation (C.10), Appendix C.

The problem (6) in the case of clipped images is rewritten as follows $\forall k \in \{1,\cdots,K\}$, $\forall i \in \{1,\dots,n_k\}$

$$\begin{cases} \tilde{\mathcal{H}}_0 &= \Big\{\tilde{z}_{k,i} \sim \tilde{P}_{\boldsymbol{\theta}_k},\ \boldsymbol{\theta}_k \in \mathbb{R}^3\Big\} \\ \tilde{\mathcal{H}}_1 &= \Big\{\tilde{z}_{k,i} \sim \tilde{Q}_{R,\boldsymbol{\theta}_k},\ \boldsymbol{\theta}_k \in \mathbb{R}^3, \forall R \in (0,1]\Big\} \end{cases} \tag{33}$$

By following the same procedure in Section 3, the GLRT for the clipped image $\tilde{\mathbf{Z}}$ is designed as follows

$$\widehat{\tilde{\delta}}^{\star} = \begin{cases} \tilde{\mathcal{H}}_0 & \text{if} \quad \widehat{\tilde{\Lambda}}^{\star}(\tilde{\mathbf{Z}}) < \widehat{\tilde{\tau}}^{\star} \\ \tilde{\mathcal{H}}_1 & \text{if} \quad \widehat{\tilde{\Lambda}}^{\star}(\tilde{\mathbf{Z}}) \geq \widehat{\tilde{\tau}}^{\star} \end{cases} \tag{34}$$

where the normalized GLR $\widehat{\tilde{\Lambda}}^{\star}(\tilde{\mathbf{Z}})$ is defined by

$$\widehat{\tilde{\Lambda}}^{\star}(\tilde{\mathbf{Z}}) = \sum_{k=1}^{K}\sum_{i=1}^{n_k} \frac{1}{\hat{\sigma}_k^2\sqrt{\hat{\tilde{v}}_0}}(\tilde{z}_{k,i} - \bar{\tilde{z}}_{k,i})(\tilde{z}_{k,i} - \hat{\mu}_k) \tag{35}$$

Similarly, the fact that the normalized GLR $\widehat{\tilde{\Lambda}}^{\star}(\tilde{\mathbf{Z}})$ follows the standard Gaussian distribution under hypothesis $\tilde{\mathcal{H}}_0$ enables to set the decision threshold independently of the image content. Under hypothesis $\tilde{\mathcal{H}}_1$, the normalized GLR $\widehat{\tilde{\Lambda}}^{\star}(\tilde{\mathbf{Z}})$ follows the Gaussian distribution with mean $\frac{\tilde{m}_1}{\sqrt{\tilde{v}_0}}$ and variance $\frac{\tilde{v}_1}{\tilde{v}_0}$ where

$$\tilde{v}_0 = \sum_{k=1}^{K} \frac{n_k\sigma_k^2\Big(1 + d_{0,k}D_{0,k} - d_{l,k}D_{l,k}\Big) + \frac{1}{\mathcal{F}(\mu_k)}}{\sigma_k^4 + \frac{a^2}{n_k\mathcal{F}(\mu_k)}} \tag{36}$$

$$\tilde{m}_1 = \frac{R}{2}\sum_{k=1}^{K}\frac{n_k}{\sigma_k^2} \tag{37}$$

$$\tilde{v}_1 = \sum_{k=1}^{K}\Bigg[\frac{n_k\sigma_k^2\Big(1 + d_{0,k}D_{0,k} - d_{l,k}D_{l,k}\Big) + \frac{1}{\mathcal{F}(\mu_k)} + \frac{n_kR}{2}}{\sigma_k^4 + \frac{a^2}{n_k\mathcal{F}(\mu_k)}} - \frac{n_kR^2}{4\sigma_k^4}\Bigg]. \tag{38}$$

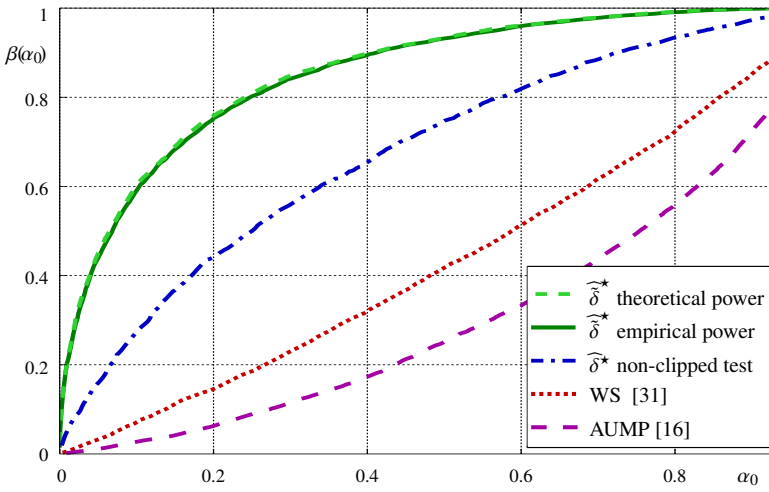

Figure 3: Detection performance on a clipped image: simulated data.

The terms $d_{0,k}$, $d_{l,k}$, $D_{0,k}$, $D_{l,k}$ are related to the clipping phenomenon and, hence, to the limited range of quantizer $\{0,\dots,l\}$. For each segment S_k, these terms are defined in Equations (C.2) and (C.5) by:

$$d_{0,k} = -\frac{\mu_k}{\sigma_k} \tag{39}$$

$$d_{l,k} = \frac{l - \mu_k}{\sigma_k} \tag{40}$$

$$D_{0,k} = \frac{\phi(d_{0,k})}{\Phi(d_{l,k}) - \Phi(d_{0,k})} \tag{41}$$

$$D_{l,k} = \frac{\phi(d_{l,k})}{\Phi(d_{l,k}) - \Phi(d_{0,k})}. \tag{42}$$

Therefore, we obtain the decision threshold and the power function of the test $\widehat{\tilde{\delta}}^{\star}$

Theorem 3. *For any clipped natural image whose parameters are unknown, the decision threshold of the test $\widehat{\tilde{\delta}}^{\star}$ based on the decision function $\widehat{\tilde{\Lambda}}^{\star}(\mathbf{Z})$ is given by :*

$$\widehat{\tilde{\tau}}^{\star} = \Phi^{-1}(1 - \alpha_0). \tag{43}$$

Theorem 4. *The power function of the test $\widehat{\tilde{\delta}}^{\star}$ is given by :*

$$\beta_{\widehat{\tilde{\delta}}^{\star}} = 1 - \Phi\left(\frac{\widehat{\tilde{\tau}}^{\star}\sqrt{\tilde{v}_0} - \tilde{m}_1}{\sqrt{\tilde{v}_1}}\right). \tag{44}$$

It can be noted that when the clipping phenomenon does not happen in a natural raw image, the power function $\beta_{\widehat{\tilde{\delta}}^{\star}}$ tends to $\beta_{\widehat{\delta}^{\star}}$. In other words, $\beta_{\widehat{\delta}^{\star}}$ can be regarded as a lower bound of $\beta_{\widehat{\tilde{\delta}}^{\star}}$ when not taking into account the clipping.

Remark 1. *For clarity, the present paper focuses on uncompressed raw images. The most challenging part when extending this work to other image formats is the impact of post-acquisition enhancement and compression processes, which is beyond the scope of this paper (see [48] for a detailed study*

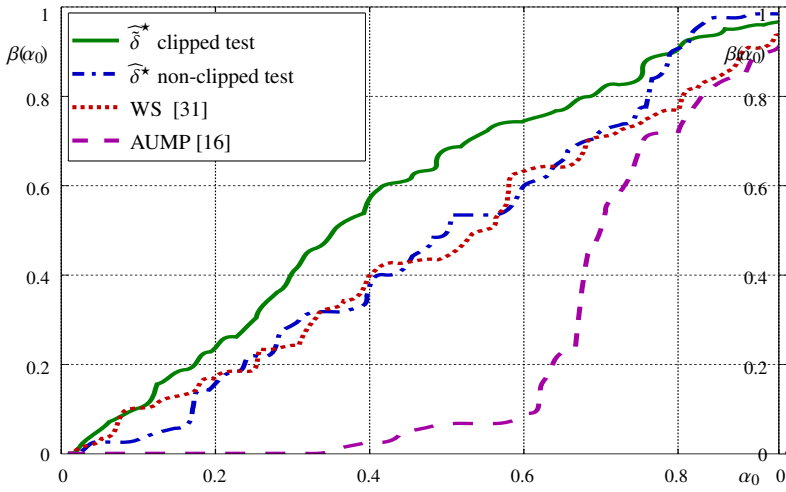

Figure 4: Detection performance on clipped real images for embedding rate $R = 0.2$.

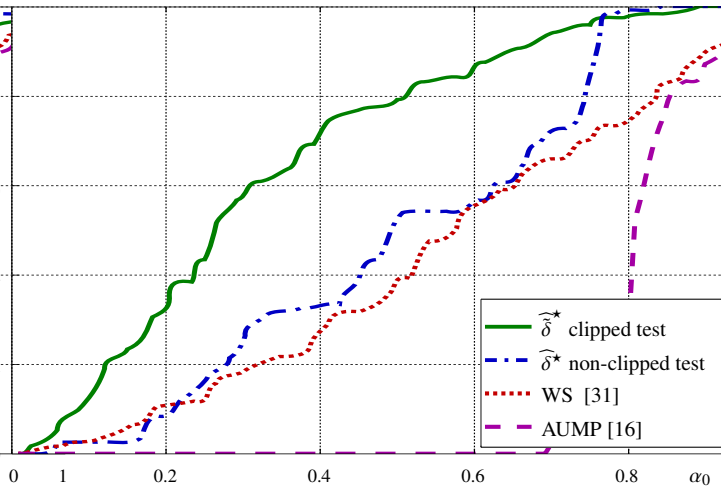

Figure 5: Detection performance on clipped real images for embedding rate $R = 0.4$.

of image acquisition pipeline). As described in [48], due to non-linear processes (white-balance, gamma-correction, etc.) the Gaussian distribution does not remain relevant to model a pixel of TIFF or JPEG image. Moreover, the spatial correlation caused by the demosaicing may lead to inaccurate noise model parameters. These difficulties need to be taken into account carefully in further research.

5. Numerical Results

5.1. Simulated Image

The detection performance of the proposed tests $\widehat{\delta}^{\star}$ and $\widehat{\tilde{\delta}}^{\star}$ is first theoretically studied on simulated data; the data is actually the 8-bit image of size 512×512 used in [20]. The heteroscedastic noise model (1) is characterized by two parameters $a = 0.0115$ and $b = 0.0002$. These values correspond to the *Nikon D70* camera with ISO 200 estimated from the Dresden image database [26]. The number of segments K is set to the number of grayscale levels, $K = 2^8$ and the embedding rate R is set at 0.05. Potentially, a large number of detectors could be compared with the proposed tests. In the present paper, it is proposed to include in the comparison the revised version of WS [31, Eq. (2)], because it is well-known for a good detection performance. In addition it is also proposed to compare the proposed tests with the AUMP detector [16] which is based on the pioneer works [6, 51] on the application of hypothesis testing theory with local regression model for estimation of unknown parameters. Moreover, this comparison particularly emphasize the crucial importance of taking into account both the heteroscedastic noise model and the clipping phenomenon. The simulation is performed with 5000 repetitions. The Figure 2 and Figure 3 illustrate respectively the detection performance of the test $\widehat{\delta}^{\star}$ and $\widehat{\tilde{\delta}}^{\star}$. The empirical power is identical to the theoretical one, which shows no loss of optimality of the proposed tests. The Figure 3 also shows that the AUMP test can not tolerate the impact of the clipping phenomenon.

5.2. Real Image Database

Experiments on real image database are then conducted to highlight the relevance of the proposed test. Clipped real images, which are captured by the *Nikon D300* camera, are provided at `http://www.cs.tut.fi/~foi/sensornoise`. Prior to our experiments, every raw image was converted to an uncompressed image format using Dcraw (with parameters -D - T -4 -j -v). Only the red color channel is used in experiments. The camera parameters are estimated on each image and used in the proposed tests. The raw images are standardized to the size of 512×512 by cropping. The state-of-the-art denoising filter for heteroscedastic noise [18] is exploited in the experiments. The experiments are performed by changing the embedding rate $R \in \{0.2, 0.4, 0.6\}$. The detection performance for each embedding rate is illustrated in Figure 4, 5 and 6. It should be noted that for the LSB replacement scheme as an insecure steganography, the embedding rate is often set to lower than 0.1. However, if the secret data is embedded in a clipped image, the higher the embedding rate is, the more probably the cover image is accepted as the stego image for WS [31] and AUMP [16]. Meanwhile, the proposed test $\widehat{\tilde{\delta}}^{\star}$ remains ensuring a high detection probability of hidden data.

To show the application on large a large image data base, experiments are conducted on 5000 raw images from the BOSS base [2]. In fact, the *Canon EOS Digital Rebel XSi* digital camera is not considered in the experiments because its images are coded with 16 bits. A change in the LSB plan is extremely negligible compared with noise level. For the same reason, raw images with very high ISO sensitivity (e.g. 1600, 3200) are also excluded from experimental dataset. On the one hand, the camera parameters $(\hat{a}, \hat{b})$ are estimated on each image. On the other hand, the parameters (a, b) are obtained by averaging the previously estimates obtained from each image with the same ISO sensitivity per camera model. The power function of the test $\widehat{\delta}^{\star}$ and $\widehat{\tilde{\delta}}^{\star}$ is drawn using the fixed camera parameters (a, b) and the estimated $(\hat{a}, \hat{b})$. The Figure 7 and Figure 8 shows the detection performance of all the detectors on 12-bit raw images

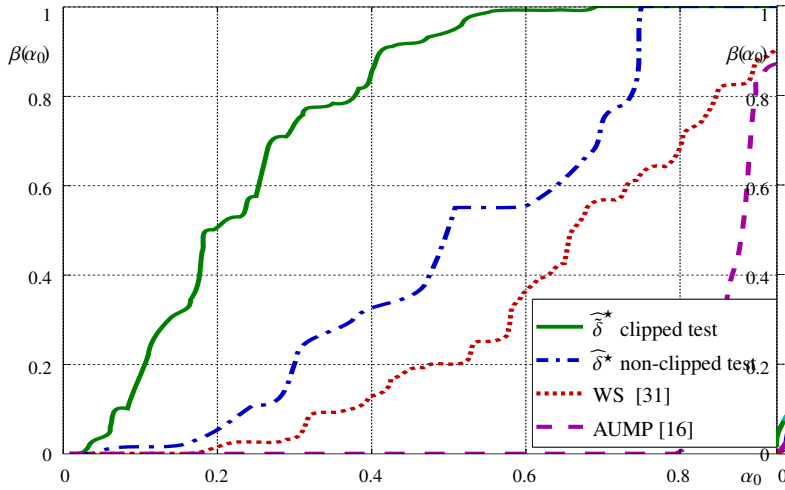

Figure 6: Detection performance on clipped real images for embedding rate $R = 0.6$.

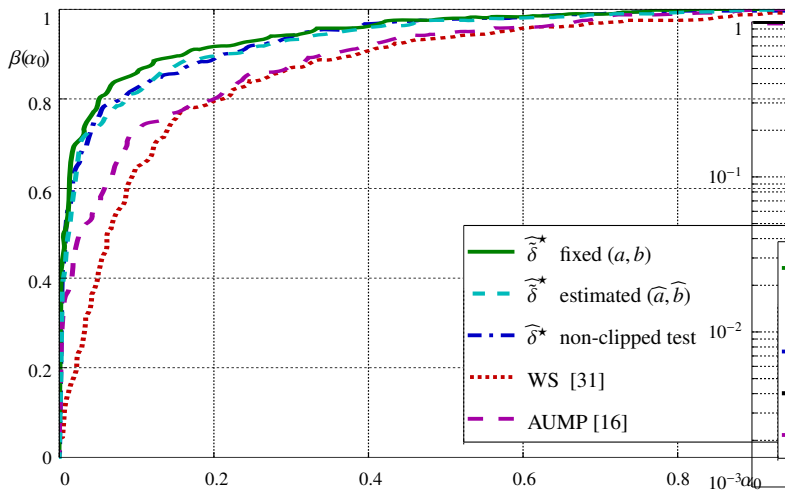

Figure 7: Detection performance on 12-bit images of Canon 400D with ISO 100 from BOSS database [2] for embedding rate $R = 0.05$.

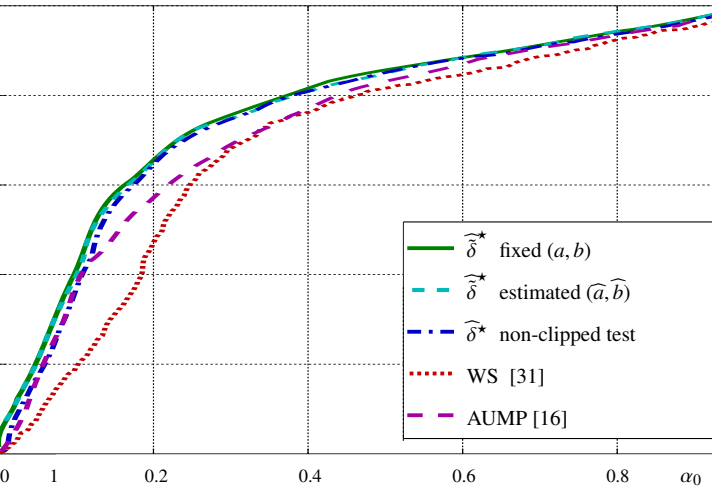

Figure 8: Detection performance on 5000 images from BOSS database [2] for embedding rate $R = 0.05$.

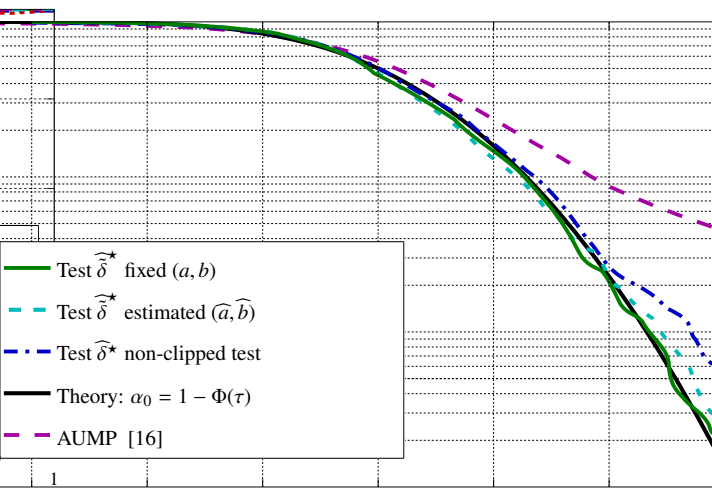

Figure 9: Empirical false-alarm probability, from real images of BOSS database [2] plotted as a function of decision threshold.

of the *Canon 400D* camera with ISO 100 and 5000 raw images from the BOSS base [2]. It is crucial to note that the clipping phenomenon seldom happens in those images. The detection performance of the AUMP and WS detectors is consequently much better than the previous results on the *Nikon D300* camera. Moreover, it is worth noting that there is no great difference between the test $\widetilde{\delta}^\star$ with (a, b) estimated for all image with the same ISO sensitivity per camera model and with $(\hat{a}, \hat{b})$ estimated for each image. This shows that the fact of estimating the camera parameters on each image does not importantly affect the power function of the test. In other words, the prior knowledge of camera parameters may not be necessary when designing the GLRTs.

Finally, it is proposed to use the previous results on 5000 images to show the possibility of the proposed test to guarantee in practice a prescribed false-alarm probability. To this end, Figure 9 shows a comparison between theoretical and empirical false-alarm probability as a function of decision threshold $\alpha_0(\tau)$. Note that, because the WS detector aims at estimating the embedding rate and do not offers to respect a false-alarm probability, this detectors is omitted from the presented results.

The numerical results presented in Figure 9 obviously shows that the AUMP detector proposed in [16] fails in practice to meet a prescribed false-alarm probability. This is mainly due to a rather inaccurate model of natural images as well as assumptions which seldom hold true such as constant variance on each block of pixels. On the opposite, Figure 9 clearly show that the proposed methodology permits in practice to guarantee a false-alarm probability. It should be noted that, again, there is a slight difference of false-alarm probability between the test $\widetilde{\delta}^\star$ with fixed (a, b) and with $(\hat{a}, \hat{b})$ estimated for each image; hence the use of estimated parameters is also possible in practice to guarantee a false-alarm probability. Similarly, the test $\widehat{\delta}^\star$ which does not take into account the clipping phenomenon slightly fails to meet a false-alarm probability in practice, particularly for a decision threshold typically corresponding to $\alpha_0 \lesssim 1.10^{-2}$.

This is mainly due to the small number of images whose clipped pixels are not negligible. Consequently, the test $\widehat{\delta}^{\star}$ is not relevant.

6. Conclusion

This paper proposes a novel methodology to detect information hidden in the LSBs plane of a natural raw image. The hidden information detection problem is cast in the framework of hypothesis testing theory. The approach is based on the heteroscedastic noise model which can be characterized only by two parameters (a, b). The main strength of the proposed approach is the design of a GLRT to detect data hidden in clipped images such that it guarantees a prescribed false alarm rate and still ensures a high detection performance while other detectors can not tolerate the impact of clipping phenomenon and fail in practice. Further research can extend this approach to attack other steganography schemes, such as LSB matching, as well as for other image formats.

Appendix A. Statistical Distribution of $\widehat{\Lambda}(\mathbf{Z})$

Appendix A.1. Statistical distribution of $\widehat{\Lambda}(\mathbf{Z})$ under hypothesis $\mathcal{H}_0$

For brevity, let $\rho_{k,i} = \frac{1}{\hat{\sigma}_k^2}(z_{k,i} - \hat{\mu}_k)$ such that

$$\widehat{\Lambda}(z_{k,i}) = \gamma_{k,i}\rho_{k,i}. \tag{A.1}$$

It is of crucial importance to analytically calculate the two first moments of $\widehat{\Lambda}(z_{k,i})$ under each hypothesis. Based on the total probability theorem, it is easily shown that

$$\begin{aligned}\mathbb{E}_{\mathcal{H}_0}\Big[\widehat{\Lambda}(z_{k,i})\Big] &= \mathbb{P}_{\mathcal{H}_0}[\gamma_{k,i}=1]\mathbb{E}_{\mathcal{H}_0}\Big[\rho_{k,i} \mid \gamma_{k,i}=1\Big] \\ &\quad - \mathbb{P}_{\mathcal{H}_0}[\gamma_{k,i}=-1]\mathbb{E}_{\mathcal{H}_0}\Big[\rho_{k,i} \mid \gamma_{k,i}=-1\Big] \\ &= \frac{1}{2}\mathbb{E}_{\mathcal{H}_0}\Big[\rho_{k,i}\Big] - \frac{1}{2}\mathbb{E}_{\mathcal{H}_0}\Big[\rho_{k,i}\Big] \\ &= 0 \end{aligned} \tag{A.2}$$

where $\mathbb{E}_{\mathcal{H}_0}[\cdot]$ and $\mathrm{Var}_{\mathcal{H}_0}[\cdot]$ denote respectively the mathematical expectation and variance under hypothesis $\mathcal{H}_0$. In virtue of the classical Delta method [3, 44], the variance of $\widehat{\Lambda}(z_{k,i})$ is given by

$$\begin{aligned}\mathrm{Var}_{\mathcal{H}_0}\Big[\widehat{\Lambda}(z_{k,i})\Big] &= \mathbb{E}_{\mathcal{H}_0}\Big[\gamma_{k,i}^2\rho_{k,i}^2\Big] \\ &= \mathbb{E}_{\mathcal{H}_0}\Big[\rho_{k,i}^2\Big] \\ &= \mathbb{E}_{\mathcal{H}_0}\Big[\frac{(z_{k,i}-\hat{\mu}_k)^2}{\hat{\sigma}_k^4}\Big] \\ &= \frac{\mathbb{E}_{\mathcal{H}_0}[(z_{k,i}-\hat{\mu}_k)^2]}{\mathbb{E}_{\mathcal{H}_0}[\hat{\sigma}_k^4]} + o(\sigma_k^{-4}) \\ &= \frac{1+\frac{1}{n_k}}{\sigma_k^2+\frac{a^2}{n_k}} + o(\sigma_k^{-4}). \end{aligned} \tag{A.3}$$

It follows from the Lindeberg central limit theorem (CLT) [35, theorem 11.2.5] that

$$\widehat{\Lambda}(\mathbf{Z}) \xrightarrow{D} \mathcal{N}(0, v_0) \tag{A.4}$$

where v_0 is the variance of the decision function $\widehat{\Lambda}(\mathbf{Z})$

$$v_0 = \sum_{k=1}^{K} \frac{n_k+1}{\sigma_k^2+\frac{a^2}{n_k}}. \tag{A.5}$$

Appendix A.2. Statistical distribution of $\widehat{\Lambda}(\mathbf{Z})$ under hypothesis $\mathcal{H}_1$

Under hypothesis $\mathcal{H}_1$, the estimates $\hat{\mu}_k$ are obviously impacted by the insertion of the secret message. However, under assumption that the impact of quantization is negligible and the denoising operator $\mathfrak{D}$ is ideal, one can obtain an almost identical denoised image

$$\mathfrak{D}(\mathbf{C}) = \mathfrak{D}(\mathbf{S}). \tag{A.6}$$

Hence, the probability that $\mathfrak{D}(z_i)$ falls into the interval $[u_k - \frac{\Delta_k}{2}, u_k + \frac{\Delta_k}{2}[$ does not change under hypothesis $\mathcal{H}_1$. It is straightforward to derive that the impact of insertion on the estimates $\hat{\mu}_k$ is negligible.

The image $\mathbf{Z}$ contains both cover and stego-pixels under hypothesis $\mathcal{H}_1$. Hence, the law of total expectation and of total variance allow to calculate the two first moments of $\widehat{\Lambda}(z_{k,i})$. The mathematical expectation of $\widehat{\Lambda}(z_{k,i})$ is defined as

$$\begin{aligned}\mathbb{E}_{\mathcal{H}_1,R}\Big[\widehat{\Lambda}(z_{k,i})\Big] &= \frac{R}{2}\mathbb{E}_{\mathcal{H}_0}\Big[\widehat{\Lambda}(\bar{z}_{k,i})\Big] \\ &\quad + \Big(1-\frac{R}{2}\Big)\mathbb{E}_{\mathcal{H}_0}\Big[\widehat{\Lambda}(z_{k,i})\Big] \\ &= \frac{R}{2}\mathbb{E}_{\mathcal{H}_0}\Big[\widehat{\Lambda}\Big(z_{k,i}+(-1)^{z_{k,i}}\Big)\Big] \\ &= \frac{R}{2}\Bigg(\mathbb{P}_{\mathcal{H}_0}[z_{k,i}=2m]\mathbb{E}_{\mathcal{H}_0}\Big[\widehat{\Lambda}(2m+1)\Big] \\ &\quad + \mathbb{P}_{\mathcal{H}_0}[z_{k,i}=2m+1]\mathbb{E}_{\mathcal{H}_0}\Big[\widehat{\Lambda}(2m)\Big]\Bigg) \\ &= \frac{R}{2}\mathbb{E}_{\mathcal{H}_0}\Big[\frac{1}{\hat{\sigma}_k^2}\Big] \\ &= \frac{R}{2\sigma_k^2} + o(\sigma_k^{-4}). \end{aligned} \tag{A.7}$$

Similarly, a direct calculation yields to

$$\begin{aligned}\mathbb{E}_{\mathcal{H}_1,R}\Big[\widehat{\Lambda}^2(z_{k,i})\Big] &= \mathbb{E}_{\mathcal{H}_1,R}\Big[\frac{(z_{k,i}-\hat{\mu}_k)^2}{\hat{\sigma}_k^4}\Big] \\ &= \frac{R}{2}\mathbb{E}_{\mathcal{H}_0}\Big[\frac{(\bar{z}_{k,i}-\hat{\mu}_k)^2}{\hat{\sigma}_k^4}\Big] \\ &\quad + \Big(1-\frac{R}{2}\Big)\mathbb{E}_{\mathcal{H}_0}\Big[\frac{(z_{k,i}-\hat{\mu}_k)^2}{\hat{\sigma}_k^4}\Big] \\ &= \mathbb{E}_{\mathcal{H}_0}\Big[\frac{(z_{k,i}-\hat{\mu}_k)^2}{\hat{\sigma}_k^4}\Big] + \frac{R}{2}\mathbb{E}_{\mathcal{H}_0}\Big[\frac{1}{\hat{\sigma}_k^4}\Big] \\ &= \frac{1+\frac{1}{n_k}+\frac{R}{2\sigma_k^2}}{\sigma_k^2+\frac{a^2}{n_k}} + o(\sigma_k^{-4}). \end{aligned} \tag{A.8}$$

It follows that the variance of $\widehat{\Lambda}(z_{k,i})$ is defined by

$$\begin{aligned}\mathrm{Var}_{\mathcal{H}_1,R}\Big[\widehat{\Lambda}(z_{k,i})\Big] &= \mathbb{E}_{\mathcal{H}_1,R}\Big[\widehat{\Lambda}^2(z_{k,i})\Big] - \mathbb{E}^2_{\mathcal{H}_1,R}\Big[\widehat{\Lambda}(z_{k,i})\Big] \\ &= \frac{1+\frac{1}{n_k}+\frac{R}{2\sigma_k^2}}{\sigma_k^2+\frac{a^2}{n_k}} - \frac{R^2}{4\sigma_k^4} + o(\sigma_k^{-4}). \end{aligned} \tag{A.9}$$

Hence, using again the Lindeberg CLT [35, theorem 11.2.5], one obtains the statistical distribution of $\hat{\Lambda}(\mathbf{Z})$ under hypothesis $\mathcal{H}_1$

$$\widehat{\Lambda}(\mathbf{Z}) \xrightarrow{D} \mathcal{N}(m_1, v_1) \tag{A.10}$$

where m_1 and v_1 are defined by

$$m_1 = \frac{R}{2}\sum_{k=1}^{K}\frac{n_k}{\sigma_k^2} \tag{A.11}$$

$$v_1 = \sum_{k=1}^{K}\Big[\frac{n_k+1+\frac{n_kR}{2\sigma_k^2}}{\sigma_k^2+\frac{a^2}{n_k}} - \frac{R^2 n_k}{4\sigma_k^4}\Big]. \tag{A.12}$$

Appendix B. Heteroscedastic Filter for Non-Clipped Noisy Images

The paper proposes a special filter to remove the heteroscedastic noise in non-clipped raw images. Motivated by the approach in [41], [18], a homomorphic transformation is used to transform an image-signal-dependent noise into an image-signal-independent noise and stabilize pixels'variance to unity. Let denote this homomorphic transformation by the function $f : [0, l] \to \mathbb{R}^+$. As explained in [41], this function is the solution of the problem

$$f(t) = \int_{-\frac{b}{a}}^{t}\frac{1}{\sigma(x)}dx = \int_{-\frac{b}{a}}^{t}\frac{1}{\sqrt{ax+b}}dx. \tag{B.1}$$

Hence, the transformation f is defined by

$$f(t) = \begin{cases}\dfrac{2}{a}\sqrt{at+b} & t > -\dfrac{b}{a} \\ 0 & t \le -\dfrac{b}{a}.\end{cases} \tag{B.2}$$

The transformation f is monotonically increasing, which implies the inverse transformation f^{-1} exists. Consequently, the inverse transformation is defined by

$$f^{-1}(t) = \frac{a}{4}t^2 - \frac{b}{a}. \tag{B.3}$$

By using the transformation f, one obtains an image $\dot{\mathbf{Z}}$ which can be treated as an image corrupted by an additive white gaussian noise

$$\dot{z} \sim \mathcal{N}(\dot{\mu}, 1). \tag{B.4}$$

The index of pixel $\dot{z}$ is omitted for seeking simplicity. At this stage, a homoscedastic filter $\mathfrak{D}_{\mathrm{ho}}$ can be applied to remove the additive white gaussian noise. Assuming that the free-noise image is successfully recovered, the denoised value $\mathfrak{D}_{\mathrm{ho}}(\dot{z})$ can be treated as the expected value of $\dot{z}$

$$\mathfrak{D}_{\mathrm{ho}}(\dot{z}) = \mathbb{E}[\dot{z}] = \dot{\mu}. \tag{B.5}$$

However, due to the non-linearity of f, the inverse transformation of the denoised value $\mathfrak{D}_{\mathrm{ho}}(\dot{z})$ can not be treated as the expected value of the pixel z

$$f^{-1}(\mathfrak{D}_{\mathrm{ho}}(\dot{z})) = f^{-1}(\dot{\mu}) \neq \mathbb{E}[z] = \mu. \tag{B.6}$$

This results in a systematic estimation bias, which needs to be corrected.

Let the function h defined by :

$$h : f(\mu) \mapsto \dot{\mu} = h(f(\mu)).$$

By using the Taylor series expansion of $f(z)$ around $z = \mu$, one obtains :

$$f(z) = f(\mu) + \sum_{n=1}^{\infty}\frac{(z-\mu)^n}{n!}f^{(n)}(\mu) \tag{B.7}$$

where $f^{(n)}$ is the n-th derivative of f. The recursive formula of $f^{(n)}$ is given by

$$f^{(n)}(t) = \Big(-\frac{a}{2}\Big)^{n-1}(2n-3)!!\,(at+b)^{-n+\frac{1}{2}}, \quad \forall n \ge 2 \tag{B.8}$$

where $n!!$ denotes the double factorial of n. Taking expectations on both sides of (B.7), the calculation shows that

$$\dot{\mu} = \mathbb{E}[f(z)] = f(\mu) - \sum_{k=1}^{\infty}\Big(\frac{a}{2}\Big)^{2k-1}\frac{(4k-3)!!}{(2k)!!}\frac{1}{\sigma^{2k-1}}. \tag{B.9}$$

Replacing $t = f(\mu)$ in (B.9), it follows that

$$\dot{\mu} = t - \sum_{k=1}^{\infty}\frac{(4k-3)!!}{(2k)!!}\frac{1}{t^{2k-1}}. \tag{B.10}$$

Therefore, the function h can be defined by

$$h(t) = t - \sum_{k=1}^{\infty}\frac{(4k-3)!!}{(2k)!!}\frac{1}{t^{2k-1}}. \tag{B.11}$$

The first derivative of h is positive, hence the function h is monotonically increasing on $\mathbb{R}^+$. This ensures the invertibility of h. Consequently, the noisy raw pixel is ideally mapped to its denoised value by the relation

$$\mu = f^{-1}(h^{-1}(\mathfrak{D}_{\mathrm{ho}}(f(z)))). \tag{B.12}$$

Due to (B.12), the special denoising filter for heteroscedastic noise in non-clipped images can be defined by

$$\mathfrak{D} = f \circ \mathfrak{D}_{\mathrm{ho}} \circ h^{-1} \circ f^{-1}. \tag{B.13}$$

Appendix C. ML Estimation of Parameters in Truncated Gaussian Data

This section proposes a ML estimation approach of parameters in truncated Gaussian data. Without loss of generality, it is assumed that a random variable X which follows the Gaussian distribution takes value on the interval $(0, l)$. The data that takes value at the boundary 0 and l is excluded. The such truncated Gaussian distribution is denoted by $\mathcal{N}_{(0,l)}(\mu, \sigma^2)$ where

$\sigma^2 = a\mu + b$. This distribution also takes advantage of the heteroscedastic relation of the mathematical expectation and variance. Suppose that the camera parameters are known in advance, the truncated Gaussian distribution $\mathcal{N}_{(0,l)}(\mu, \sigma^2)$ is hence characterized by the only one parameter μ. The pdf of X is simplified from (28)

$$f_X(x) = \frac{1}{\Phi(d_l) - \Phi(d_0)} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big[- \frac{(x-\mu)^2}{2\sigma^2} \Big] \qquad \text{(C.1)}$$

where

$$d_0 = \frac{-\mu}{\sigma} \quad \text{and} \quad d_l = \frac{l-\mu}{\sigma}. \qquad \text{(C.2)}$$

It must be noted that now μ is not the expected value of X. From [29], the expected value and variance of X are given by

$$\mathbb{E}[X] = \mu + (D_0 - D_l)\sigma \qquad \text{(C.3)}$$

$$\mathrm{Var}[X] = \sigma^2\Big[1 + d_0 D_0 - d_l D_l - (D_0 - D_l)^2\Big] \qquad \text{(C.4)}$$

where

$$D_0 = \frac{\phi(d_0)}{\Phi(d_l) - \Phi(d_0)} \quad \text{and} \quad D_l = \frac{\phi(d_l)}{\Phi(d_l) - \Phi(d_0)}. \qquad \text{(C.5)}$$

Suppose now a vector of n independent random variables $\mathbf{X} = \{X_i\}_{i=1}^n$ following the truncated Gaussian distribution $\mathcal{N}_{(0,l)}(\mu, \sigma^2)$

$$X_i \sim \mathcal{N}_{(0,l)}(\mu, \sigma^2). \qquad \text{(C.6)}$$

By definition, the ML estimate $\hat{\mu}$ is defined as the solution which maximizes the log likelihood function of the samples $\mathbf{X}$

$$\hat{\mu} = \arg\max_{\mu} L(\mathbf{X}|\mu) = \arg\max_{\mu} \sum_{i=1}^{n} L(X_i|\mu) \qquad \text{(C.7)}$$

where

$$L(X_i|\mu) = -\log\Big[\Phi(d_l) - \Phi(d_0)\Big] - \log(\sqrt{2\pi\sigma^2}) - \frac{(X_i - \mu)^2}{2\sigma^2}. \qquad \text{(C.8)}$$

Unfortunately, the ML estimate $\hat{\mu}$ can not be analytically given. This problem has been studied in [11] by solving a system of two differential equations based on a modified Newton-Raphson method [50]. In this paper, the problem (C.7) is numerically solved using the Nelder-Mead optimization method [36] and the sample mean $\bar{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i$ as initial solution. From [44], the ML estimates is asymptotically unbiased and distributed as Gaussian

$$\hat{\mu} \sim \mathcal{N}\Big(\mu, \frac{1}{n\mathcal{F}(\mu)}\Big) \qquad \text{(C.9)}$$

where $\mathcal{F}(\mu)$ represents the Fisher information

$$\begin{aligned}\mathcal{F}(\mu) &= -\mathbb{E}\left[\frac{\partial^2 L(X_i|\mu)}{\partial\mu^2}\right] \\ &= \frac{\partial^2 d_l}{\partial\mu^2} D_l + \frac{\partial d_l}{\partial\mu}\frac{\partial D_l}{\partial\mu} - \frac{\partial^2 d_0}{\partial\mu^2} D_0 - \frac{\partial d_0}{\partial\mu}\frac{\partial D_0}{\partial\mu} \\ &\quad - \frac{a^2\sigma^2 - 2b^2}{2\sigma^6} + \frac{a(a^2\mathbb{E}[X_i^2] + 2b\mathbb{E}[X_i])}{\sigma^6}. \end{aligned} \qquad \text{(C.10)}$$

References

[1] H. H. Barrett and K. J. Myers. *Foundations of Image Science*. Wiley, Newyork, 2004.

[2] P. Bas, T. Filler, and T. Pevný. Break our steganographic system — the ins and outs of organizing boss. In T. Filler, editor, *Information Hiding, 13th International Workshop*, 2011.

[3] P. Billingsley. *Probability and Measure*. Ser. Wiley Series in Probability and Mathematical Statistics, Newyork, 1995.

[4] D. Birkes. Generalized likelihood ratio tests and uniformly most powerful tests. *The American Statistician*, 44(2):163 – 166, 1990.

[5] R. Böhme. *Advanced Statistical Steganalysis*. Springer, New York, 2010.

[6] R. Cogranne, C. Zitzmann, L. Fillatre, F. Retraint, I. Nikiforov, and P. Cornu. Statistical decision by using quantized observations. In *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*, 1210 – 1214, 2011.

[7] R. Cogranne and F. Retraint. Application of hypothesis testing theory for optimal detection of lsb matching data hiding. In *Elsevier Signal Processing*, 93(7):1724 – 1737, 2013.

[8] R. Cogranne and F. Retraint. An asymptotically uniformly most powerful test for lsb matching detection. *Information Forensics and Security, IEEE Transactions on*, 8(3):464 – 476, March 2013.

[9] R. Cogranne, C. Zitzmann, L. Fillatre, I. Nikiforov, F. Retraint, and P. Cornu. A cover image model for reliable steganalysis. In *Information Hiding, 13th International Workshop*, Lecture Notes in Computer Science, pages 178 – 192, Prague, Czech Republic, May 18–20, 2011. LNCS vol.6958, Springer-Verlag, New York.

[10] R. Cogranne, C. Zitzmann, L. Fillatre, I. Nikiforov, F. Retraint, and P. Cornu. Reliable detection of hidden information based on a non-linear local model. In *Statistical Signal Processing, Proc. of IEEE Workshop on*, pages 493 – 496, 28-30 June 2011.

[11] A. Cohen. Estimating the mean and variance of normal populations from singly truncated and doubly truncated samples. *Ann. Math. Statist.*, 21(4):557 – 569, 1950.

[12] I. Cox, M. Miller, J. Bloom, J. Fridrich, and T. Kalker. *Digital Watermarking and Steganography*. Morgan Kaufmann, San Francisco, CA, 2007.

[13] O. Dabeer, K. Sullivan, U. Madhow, S. Chandrasekaran, and B. Manjunath. Detection of hiding in the least significant bit. *Signal Processing, IEEE Transactions on*, 52(10):3046 – 3058, October 2004.

[14] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *Image Processing, IEEE Transactions on*, 16(8):2080 – 2095, August 2007.

[15] S. Dumitrescu, X. Wu, and W. Zhe. Detection of lsb steganography via sample pair analysis. *IEEE Trans. Signal Process.*, 51(6):1995 – 2007, 2003.

[16] L. Fillatre. Adaptive steganalysis of least significant bit replacement in grayscale natural images. *Signal Processing, IEEE Transactions on*, 60(2):556 – 569, February 2012.

[17] L. Fillatre, I. Nikiforov, and F. Retraint. ϵ-optimal non-bayesian anomaly detection for parametric tomography. *IEEE Trans. Image Process.*, 17(11):1985 – 1999, 2008.

[18] A. Foi. Clipped noisy images: heteroskedastic modeling and practical denoising. *Signal Processing*, 89(12):2609 – 2629, 2009.

[19] A. Foi, V. Katkovnik, and K. Egiazarian. Signal-dependent noise removal in pointwise shape-adaptive DCT domain with locally adaptive variance. In *Proc. EUSIPCO 2007*, Poznan, September 2007.

[20] A. Foi, M. Trimeche, V. Katkovnik, and K. Egiazarian. Practical poissonian-gaussian noise modeling and fitting for single-image raw-data. *IEEE Trans. Image Process.*, 17(10):1737 – 1754, October 2008.

[21] M. Fouladirad and I. Nikiforov. Optimal statistical fault detection with nuisance parameters. *Automatica*, 41:1157 – 1171, 2005.

[22] J. Fridrich. *Steganography in Digital Media : Principles, Algorithms, and Applications*. Cambridge Univ. Press, Newyork, 2009.

[23] J. Fridrich and M. Goljan. On estimation of secret message length in LSB steganography in spatial domain. In *Security, Steganography and Watermarking of Multimedia Contents VI. Proc. of the SPIE*, volume 5306, 2004.

[24] J. Fridrich, M. Goljan, and R. Du. Reliable detection of lsb steganography in color and grayscale images. *IEEE Multimedia*, 8:22 – 28, 2001.

[25] J. Fridrich and J. Kodovskỳ. Steganalysis of LSB replacement using

parity-aware features. In *Information Hiding Conference,* i*n press*. Lecture Notes in Computer Science, Springer, 15 - 18 May 2012.

[26] T. Gloe and R. Bohme. The 'dresden image database' for benchmarking digital image forensics. *Proc. ACM SAC*, 2:1585 – 1591, 2010.

[27] G. E. Healey and R. Kondepudy. Radiometric ccd camera calibration and noise estimation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 16:267 – 276, March 1994.

[28] K. Hirakawa and T. Parks. Image denoising for signal-dependent noise. In *Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05). IEEE International Conference on*, volume 2, pages 29 – 32, 18-23, 2005.

[29] N. L. Johnson, S. Kotz, and N. Balakrishnan. *Continuous univariate distributions*. Wiley, 2rd edition, 1994.

[30] A. Ker. Fourth-order structural steganalysis and analysis of cover assumptions. In *Proc. SPIE 6072*, pages 25 – 38, 2006.

[31] A. Ker and R. Böhme. Revisiting weighted stego-image steganalysis. In *Security, Steganography and Watermarking of Multimedia Contents X. Proc. of the SPIE*, pages 501 – 517, 2008.

[32] J. Kodovsky, J. Fridrich, and V. Holub. Ensemble classifiers for steganalysis of digital media. *Information Forensics and Security, IEEE Transactions on*, 7(2):432 – 444, 2012.

[33] L. Le Cam. *Asymptotics Methods in Statistical Decision Theory*. Series in Statistics, Springer, Newyork, 1986.

[34] K. Lee et al. Generalized category attack : Improving histogram-based attack on JPEG LSB embedding. In *Proc. 9th Int. Workshop Information Hiding*, volume 4567, pages 378 – 392. Springer, 2007.

[35] E. L. Lehmann and J. P. Romano. *Testing Statistical Hypotheses*. Springer, Newyork, 3rd edition, 2005.

[36] J. Nelder and R. Mead. A simplex method for function minimization. *The Computer Journal*, 7:308 – 313, 1965.

[37] A. Nissar and A. Mir. Classification of steganalysis techniques : A study. *Digit. Signal Process.*, 20(6):1758 – 1770, 2010.

[38] T. Pevny, P. Bas, and J. Fridrich. Steganalysis by subtractive pixel adjacency matrix. *information Forensics and Security, IEEE Transactions on*, 5(2):215 – 224, 2010.

[39] B. Porat and B. Friedlander. On the generalized likelihood ratio test for a class of nonlinear detection problems. *IEEE Trans. Signal Process.*, 41(11):3186 – 3190, 1993.

[40] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli. Image denoising using scale mixtures of gaussians in the wavelet domain. *Image Processing, IEEE Transactions on*, 12(11):1338 – 1351, November 2003.

[41] P. Prucnal and B. Saleh. Transformation of image-signal-dependent noise into image-signal-independent noise. *Optics letters*, 6(7):316 – 318, July 1981.

[42] L. Scharf and B. Friedlander. Matched subspace detectors. *IEEE Trans. Signal Process.*, 42(8):2146 – 2157, 1994.

[43] C. Scott. Performance measures for Neyman-Pearson classification. *Information Theory, IEEE Transactions on*, 53(8):2852 – 2863, aug. 2007.

[44] A. Stuart and J. Ord. *Kendall's Advanced Theory of Statistics*, volume 1. Arnold, 6rd edition, 1994.

[45] K. Sullivan, U. Madhow, S. Chandrasekaran, and B. Manjunath. Steganalysis for markov cover data with applications to images. *Information Forensics and Security, IEEE Transactions on*, 1(2):275 – 287, June 2006.

[46] T. H. Thai, R. Cogranne, and F. Retraint. Camera model identification Based on hypothesis testing theory. *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European*, pages 1747 – 1751, August 2012.

[47] T. H. Thai, R. Cogranne, and F. Retraint. Camera Model Identification based on the Heteroscedastic Noise Model. be accepted with minor revisions to *Image Processing, IEEE Transactions on*, 2012.

[48] T. H. Thai, F. Retraint, and R. Cogranne. Statistical Model of Natural Images. In *IEEE International Conference on Image Processing, ICIP*, pages 2525 – 2528, September 2012.

[49] Y. Wang and P. Moulin. Optimized feature extraction for learning-based image steganalysis. *Information Forensics and Security, IEEE Transactions on*, 2(1):31 – 45, March 2007.

[50] E. Whitaker and G. Robinson. *The Calculus of Observations*. Blackie and Son, Ltd., London and Glasgow, 2rd edition, 1929.

[51] C. Zitzmann, R. Cogranne, F. Retraint, I. Nikiforov, L. Fillatre, and P. Cornu. Statistical decision methods in hidden information detection. In *Information Hiding*, volume LNCS 6958, pages 163 – 177, Prague, Czech Republic, 2011.