Statistical Detection of LSB Matching Using Hypothesis Testing Theory

Rémi Cogranne, Cathel Zitzmann, Florent Retraint, Igor Nikiforov, Lionel Fillatre and Philippe Cornu *

ICD - LM2S - Université de Technologie de Troyes - UMR STMR CNRS
12, rue Marie Curie - B.P. 2060 - 10010 Troyes cedex - France
E-mail : name.surname@utt.fr

Abstract. This paper investigates the detection of information hidden by the Least Significant Bit (LSB) matching scheme. In a theoretical context of known image media parameters, two important results are presented. First, the use of hypothesis testing theory allows us to design the Most Powerful (MP) test. Second, a study of the MP test gives us the opportunity to analytically calculate its statistical performance in order to warrant a given probability of false-alarm. In practice when detecting LSB matching, the unknown image parameters have to be estimated. Based on the local estimator used in the Weighted Stego-image (WS) detector, a practical test is presented. A numerical comparison with state-of-the-art detectors shows the good performance of the proposed tests and highlights the relevance of the proposed methodology.

1 Introduction and Contributions.

Steganography and steganalysis form a cat-and-mouse game. On the one hand, steganography aims at hiding the very presence of a secret message by hiding it within an innocuous cover medium. On the other hand, the goal of steganalysis (in the wide sense) is to obtain any information about the potential steganographic system from an unknown medium. Usually, steganalysis focuses on exposing the existence of a hidden message in an inspected medium.
Many steganographic tools are nowadays easily available on the Internet making steganography within the reach of anyone, for legitimate or malicious usage. It is thus crucial for security forces to be able to reliably detect steganographic content among a (possibly very large) set of media files. In this operational context, the detection of a rather simple but most commonly found stegosystem seems more important than the detection of a very complex but rarely encountered

* This work was partially supported by National Agency for Research (ANR) through ANR-CSOSG Program (Project ANR-07-SECU-004). 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.

stegosystem. The vast majority of downloadable steganographic tools insert the secret information in the LSB plane. Consequently, substantial progress has recently been made in the detection of such steganographic algorithms, namely LSB replacement and LSB matching, also known as LSB ± 1 embedding (see [11, 15, 1] and the references therein). However, the steganalysis of LSB matching remains much harder than the steganalysis of LSB replacement. Indeed, if LSB matching is used instead of LSB replacement, the detection power of state-of-the-art detectors is significantly lower [25, 5].

The recently proposed steganalyzers dedicated to LSB matching can be roughly divided into two categories. On the one hand, most of the latest detectors are based on supervised machine learning methods and use targeted [6, 4] or universal features [17, 23]. As in all applications of machine learning, the theoretical calculation of error probabilities remains an open problem [24]. On the other hand, the authors of [18] observed that LSB matching acts as a low-pass filter on the image Histogram Characteristic Function (HCF). This pioneering work lead to an entire family of histogram-based detectors [19, 25].

In the operational context described above, the proposed steganalyzer must be immediately applicable without any training or tuning phase. For this reason, the use of a machine learning based detector is hardly possible. Moreover, the most important challenge for the steganalyst is to provide detection algorithms with an analytical expression for the false-alarm and missed-detection probabilities without which the "uncertainty" of the result can not be "measured." The proposed LSB matching steganalyzers are certainly very interesting and efficient, but these *ad hoc* algorithms have been designed with a very limited exploitation of statistical cover models and hypothesis testing theory. Hence, a few theoretical results exist and the only solution to measure their statistical performance is the simulation on large databases.

Alternatively, the first step in the direction of hypothesis testing has been made in [12, 8, 9] for LSB replacement to design a statistical test with known statistical properties. In the present paper, this statistical approach is extended to the case of detecting LSB matching. More precisely, the goal of this paper is threefold:

1. Define the most powerful (MP) test in the theoretical case when the cover image parameters are known, namely the expectation and noise variance of each pixel.
2. Analytically calculate the statistical performance of the MP test in terms of the false-alarm and missed-detection probabilities. More importantly, this result allows us to highlight the impact of the noise variance and quantization on the test performance [9].
3. Design a practical efficient implementation of this test based on a simple local estimation of expectation and variance of each pixel.

The paper is organized as follows. The problem of LSB matching steganalysis is casted within the framework of hypothesis testing in Section 2. Following the Neyman-Pearson approach, the MP Likelihood Ratio Test (LRT) is presented in Section 3 and its statistical performance is calculated in Section 4. Finally, the

proposed practical implementation of the Generalized LRT (GLRT) is presented in Section 5. To show the relevance of the proposed approach, numerical results on large natural image databases are shown in Section 6. Section 7 concludes the paper.

2 Detection of LSB Matching Problem Statement.

This paper mainly focuses on natural images but the extension of the presented results to any kind of digital media is immediate. Hence, the column vector $\boldsymbol{C} = (c_1, \ldots, c_N)^T$ represents in this paper a cover image of $N = N_x \times N_y$ grayscale pixels. The set of grayscale levels is denoted $\mathcal{Z} = \{0; \ldots; 2^{B-1}\}$ as pixels values are usually unsigned integers encoded with B bits. Each cover pixel c_n results from the quantization:

$$c_n = Q(y_n), \tag{1}$$

where $y_n \in \mathbb{R}^+$ denotes the raw pixel intensity recorded by the camera and Q represents the uniform quantization with a unitary step:

$$Q(x) = k \Leftrightarrow x \in [k - 1/2\,;\, k + 1/2[.$$

Seeking simplicity, it is assumed in this paper that the saturation effect is absent, *i.e.* the probability of excessing the quantizer boundaries $-1/2$ and $2^{B-1}+1/2$ is negligible. Indeed, taking into account the under or over-exposed pixels is rather simple but requires a much more complicated notation.
The recorded pixel value can be decomposed as [13, 7]:

$$y_n = \theta_n + \xi_n, \tag{2}$$

where θ_n is a deterministic parameter corresponding to the mathematical expectation of y_n and ξ_n is a random variable representing all the noise corrupting the cover image during acquisition. As described in [13], ξ_n is accurately modeled as a realization of a zero-mean Gaussian random variable $\Xi_n \sim \mathcal{N}(0, \sigma_n^2)$ whose variance σ_n^2 varies from pixel to pixel. It thus follows from (1) and (2) that c_n follows a distribution $P_{\theta_n} = P_{\theta_n,\sigma_n} = (p_{\theta_n}[0], \ldots, p_{\theta_n}[2^{B-1}])$ defined by:

$$\forall k \in \mathcal{Z}\ ,\ p_{\theta_n}[k] = \Phi\left(\frac{k + 1/2 - \theta_n}{\sigma_n}\right) - \Phi\left(\frac{k - 1/2 - \theta_n}{\sigma_n}\right), \tag{3}$$

with Φ is the standard Gaussian cumulative distribution function (cdf) defined by $\Phi(x) = \int_{-\infty}^{x} \phi(u)du$ and ϕ the standard Gaussian probability distribution function (pdf) $\phi(u) = \frac{1}{\sqrt{2\pi}} \exp(u^2/2)$. In virtue of the mean value theorem, (3) can be written as:

$$p_{\theta_n}[k] = \frac{1}{\sigma_n} \int_{k-\frac{1}{2}}^{k+\frac{1}{2}} \phi\left(\frac{u - \theta_n}{\sigma_n}\right) du = \phi\left(\frac{k - \theta_n}{\sigma_n} + \epsilon\right), \tag{4}$$

where ϵ is a (small) corrective term [26].

To statistically model stego-image pixels from (3)–(4), the two following assumptions are usually adopted [12, 14] : 1) the probability of insertion is equal for every cover pixel (independence between hidden bits and cover pixels) and 2) the message is assumed compressed and/or cyphered $\boldsymbol{M} = (m_1, \ldots, m_L)^T$ before insertion. Hence, each hidden bit m_l is drawn from a binomial distribution $\mathcal{B}(1, 1/2)$, *i.e.* m_l is either 0 or 1 with the same probability. This situation is captured by denoting

$$\forall n \in \{0, \ldots, N\}, \begin{cases} \mathbb{P}[s_n = c_n] = (1-R), \\ \mathbb{P}[s_n = c_n + \text{ins}(m_n, c_n)] = R, \end{cases} \tag{5}$$

where $\boldsymbol{S} = \{s_1, \ldots, s_N\}$ are the values of stego-image pixels, the embedding rate $R = L/N$ corresponds to the number of hidden bits per cover pixel and $\text{ins}(m_n, c_n)$ represents the value added to c_n to insert the hidden bit m_n.
The particularity of LSB matching lies in its insertion function $\text{ins} : \{0; 1\} \times \mathcal{Z} \mapsto \{-1; 0; 1\}$. Whenever the LSB of c_n is equal to m_n, *i.e.* when $\text{lsb}(c_n) = c_n \text{mod} 2 = m_n$, there is no need to change c_n, hence $\text{ins}(m_n, c_n) = 0$. On the contrary, whenever $\text{lsb}(c_n) \neq m_n$, the insertion must change the LSB of c_n, which is done by adding or subtracting 1 with the same probabilities:

$$\begin{cases} \mathbb{P}[\text{ins}(b_s, c_n) = 1 \,|\, \text{lsb}(c_n) \neq m_n] &= 1/2 \\ \mathbb{P}[\text{ins}(b_s, c_n) = -1 \,|\, \text{lsb}(c_n) \neq m_n] &= 1/2. \end{cases} \tag{6}$$

Since each hidden bit m_n follows the binomial distribution $\mathcal{B}(1, 1/2)$, a straightforward calculation finally shows that $\mathbb{P}[\text{lsb}(c_n) = m_n] = \mathbb{P}[\text{lsb}(c_n) \neq m_n] = 1/2$. Hence, as described in [18, 25, 6, 10], it follows from (5)–(6) that for all $n \in \{1, \ldots, N\}$, the pmf of the stego-pixel s_n after embedding at rate R with LSB matching is given by $Q^R_{\theta_n} = \left(q^R_{\theta_n}[0], \ldots, q^R_{\theta_n}[2^b - 1]\right)$ with $\forall k \in \mathcal{Z}$:

$$q^R_{\theta_n}[k] = \frac{R}{4}\left(p_{\theta_n}[k-1] + p_{\theta_n}[k+1]\right) + \left(1 - \frac{R}{2}\right) p_{\theta_n}[k]. \tag{7}$$

3 Likelihood Ratio Test (LRT) for two simple hypotheses.

When analyzing an unknown medium $\mathbf{Z}$ the first goal of LSB matching steganalysis is to decide between the two following hypotheses:

$$\begin{aligned} \mathcal{H}_0 &= \{z_n \sim P_{\theta_n}, \forall n \in \{1, \ldots, N\}\} \\ \text{vs} \quad \mathcal{H}_1 &= \{z_n \sim Q^R_{\theta_n}, \forall n \in \{1, \ldots, N\}\}. \end{aligned} \tag{8}$$

Let us start with the simplest case, when the embedding rate R and, for all n, the parameters θ_n and σ_n are known. In this case, the hypothesis testing problem (8) is reduced to a test between two simple hypotheses.
The goal is obviously to find a test $\delta : \mathcal{Z}^N \mapsto \{\mathcal{H}_0, \mathcal{H}_1\}$, such that hypothesis $\mathcal{H}_i$ is accepted if $\delta(\mathbf{Z}) = \mathcal{H}_i$ (see [22] for details about statistical hypothesis testing). However, as explained in the introduction, in an operational forensics

context the most important challenge is first, to warrant a prescribed (very low) false-alarm probability and second, to maximize the detection power defined by:

$$\beta_\delta = \mathbb{P}_1[\delta(\mathbf{Z}) = \mathcal{H}_1],$$

where $\mathbb{P}_i(\cdot)$ stands for the probability under hypotheses $\mathcal{H}_i$, $i = \{0; 1\}$. Therefore, let $\mathcal{K}_\alpha$ be the class of tests with an upper-bounded false-alarm probability α_0 defined by

$$\mathcal{K}_\alpha = \{\delta : \mathbb{P}_0[\delta(\mathbf{Z}) = \mathcal{H}_1] \leq \alpha_0\}. \tag{9}$$

In virtue of the Neyman-Pearson lemma, see [22, Theorem 3.2.1], the most powerful (MP) test over the class $\mathcal{K}_{\alpha_0}$ (9) is the LRT given by the following decision rule:

$$\delta_R(\mathbf{Z}) = \begin{cases} \mathcal{H}_0 & \text{if} \quad \Lambda_R(\mathbf{Z}) \leq \tau_{\alpha_0} \\ \mathcal{H}_1 & \text{if} \quad \Lambda_R(\mathbf{Z}) > \tau_{\alpha_0}, \end{cases} \tag{10}$$

where τ_{α_0} is the solution of $\mathbb{P}_0[\delta(\mathbf{Z}) > \tau_{\alpha_0}] = \alpha_0$, to insure that $\delta_R \in \mathcal{K}_{\alpha_0}$, and the likelihood ratio (LR) $\Lambda_R(\mathbf{Z})$ is given, from the statistical independence between pixels, by:

$$\Lambda_R(\mathbf{Z}) = \prod_{n=1}^{N} \Lambda_R(z_n) = \prod_{n=1}^{N} \frac{R}{4} \frac{p_{\theta_n}[z_n - 1] + p_{\theta_n}[z_n + 1]}{p_{\theta_n}[z_n]} + \left(1 - \frac{R}{2}\right). \tag{11}$$

It can be noted that $\Lambda_R(z_n)$ depends on pixel values z_n through the quantity:

$$\Lambda_2(z_n) = \frac{1}{2} \frac{p_{\theta_n}[z_n - 1] + p_{\theta_n}[z_n + 1]}{p_{\theta_n}[z_n]}, \tag{12}$$

which corresponds to the the likelihood ratio for the conceptual case of $R = 2$. In other words, Equation (12) corresponds to this test: $\mathcal{H}_0 : \{\, \mathbf{Z} \text{ is a cover medium} \,\}$ vs $\mathcal{H}_1 : \{\, \text{each pixel of } \mathbf{Z} \text{ is modified by} \pm 1 \,\}$. Indeed, considering the case $R{=}2$ permits us to clarify the present methodology, which is then extended to the more general case of $R \in]0; 1[$ in Section 4.2.

The exact expression for the LR $\Lambda_2(z_n)$ is complicated due to the corrective terms ϵ defined in (4). However, the calculation shows that these corrective terms are usually negligible, particularly when $\sigma_n > 1$. Therefore, it is proposed to neglect ϵ in order to obtain a simplified expression for the LR $\Lambda_2(z_n)$. From (4), this approximation permits us to write:

$$\begin{aligned} \frac{p_{\theta_n}[z_n - 1]}{p_{\theta_n}[z_n]} &= \exp\left(-\frac{1}{2\sigma_n^2}\right) \exp\left(\frac{\theta_n - z_n}{\sigma_n^2}\right), \\ \frac{p_{\theta_n}[z_n + 1]}{p_{\theta_n}[z_n]} &= \exp\left(-\frac{1}{2\sigma_n^2}\right) \exp\left(\frac{z_n - \theta_n}{\sigma_n^2}\right). \end{aligned} \tag{13}$$

Finally, using (13), the LR $\Lambda_2(z_n)$ can be written as:

$$\Lambda_2(z_n) = \frac{1}{4} \exp\left(\frac{-1}{2\sigma_n^2}\right) \left[\exp\left(\frac{z_n - \theta_n}{\sigma_n^2}\right) + \exp\left(\frac{\theta_n - z_n}{\sigma_n^2}\right)\right]. \tag{14}$$

The logarithm of the likelihood ratio (15) is usually preferred in order to replace the product in (11) with a sum. From (14), it immediately follows that:

$$\begin{aligned}\widetilde{\Lambda}_2(z_n) &\stackrel{\text{def.}}{=} \log\left[\exp\left(\frac{z_n-\theta_n}{\sigma_n^2}\right)+\exp\left(\frac{\theta_n-z_n}{\sigma_n^2}\right)\right] \\ &= \log\big(\Lambda_2(z_n)\big)+\log(2)+\frac{1}{2\sigma_n^2}.\end{aligned} \tag{15}$$

Again, one can note that the terms $\log(4)$ and $\frac{1}{2\sigma_n^2}$ do not depend on the true hypothesis. That is why, for the same reasons as those discussed in connection with Equation (12), these terms do not play any role in solving the detection problem (8). For the sake of clarity, these terms are thus omitted from expression (15) of the log-LR $\widetilde{\Lambda}_2(z_n)$.

4 Statistical Performance of the LR test.

4.1 Case of simple hypotheses, when $R = 2$.

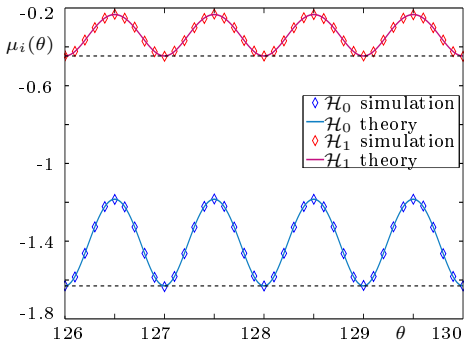
In this section it is first proposed to study the statistical performance for the case of simple hypotheses, when $R = 2$. The results are then extended to the general case of $R \in]0;1[$ in Section 4.2. To easily calculate the statistical performance of the LR test δ_R (10), the asymptotic approach is of crucial interest. Moreover, the assumption that N grows to infinity is relevant in practice due to the very large number of pixels in typical images.
For the sake of clarity, let the mean expectation and the mean variance of $\widetilde{\Lambda}_2(z_n)$ under hypotheses $\mathcal{H}_i$ be defined as follows:

$$\mu_i = \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}_i\Big[\widetilde{\Lambda}_2(z_n)\Big] \quad \text{and} \quad \sigma_i^2 = \frac{1}{N}\sum_{n=1}^{N}\mathbb{V}ar_i\Big[\widetilde{\Lambda}_2(z_n)\Big], \tag{16}$$

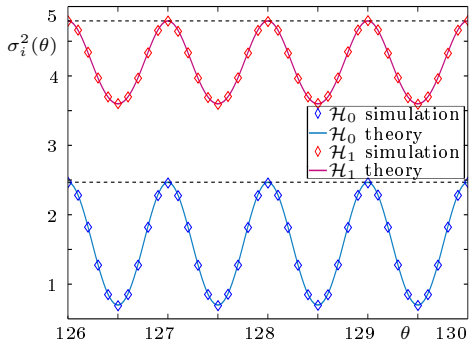
where $\mathbb{E}_i\big[\widetilde{\Lambda}_2(\mathbf{Z})\big]$ and $\mathbb{V}ar_i\big[\widetilde{\Lambda}_2(\mathbf{Z})\big]$ are respectively the expectation and the variance of $\widetilde{\Lambda}_2(z_n)$ under hypotheses $\mathcal{H}_i$, $i = \{0, 1\}$.
The test $\widetilde{\delta}_2$ associated with the "normalized" log-LR $\widetilde{\Lambda}_2(\mathbf{Z})$ is defined as:

$$\widetilde{\delta}_2 = \begin{cases}\mathcal{H}_0 & \text{if} \quad \widetilde{\Lambda}_2(\mathbf{Z}) \leq \widetilde{\tau}_{\alpha_0}, \\ \mathcal{H}_1 & \text{if} \quad \widetilde{\Lambda}_2(\mathbf{Z}) > \widetilde{\tau}_{\alpha_0}.\end{cases} \quad \text{where} \quad \widetilde{\Lambda}_2(\mathbf{Z}) \stackrel{\text{def.}}{=} \frac{\sum_{n=1}^{N}\widetilde{\Lambda}_2(z_n) - N\mu_0}{\sqrt{N\sigma_0^2}}, \tag{17}$$

It can noted that the random variables $\widetilde{\Lambda}_2(z_n)$ are assumed statistically independent and, for any $\sigma_n > 0$, have finite expectation and variance, which implies that the conditions necessary for application of the Lindeberg's central limit theorem [22, Theorem 11.2.5] are satisfied. These conditions can also be shown by using the fact that z_n are bounded because they can only take values in the set



(a) Expectations of the log-LR $\log\left(\Lambda_2(z_n)\right)$ as a function of expectation θ.



(b) Variances of the log-LR $\log\left(\Lambda_2(z_n)\right)$ as a function of expectation θ.

Fig. 1: Graphical representation of the two first moments of log-LR $\log\left(\Lambda_2(z_n)\right)$ (20) - (23). Presented results correspond to the case of i.i.d pixels with expectation $\theta_n \in [126; 130]$ and standard deviation $\sigma_n = 0.75$.

$\mathcal{Z}$. Therefore,

$$\widetilde{\Lambda}_2(\mathbf{Z}) \rightsquigarrow \begin{cases} \mathcal{N}(0\,,\,1) & \text{under} \quad \mathcal{H}_0 \\ \mathcal{N}\left(\dfrac{\sqrt{N}(\mu_2 - \mu_0)}{\sigma_0}\,,\,\dfrac{\sigma_2^2}{\sigma_0^2}\right) & \text{under} \quad \mathcal{H}_1. \end{cases} \tag{18}$$

where $\rightsquigarrow$ represents the convergence in distribution as $N \to \infty$. From Equation (18), a short algebra establishes the following theorem.

Theorem 1. *For any given probability of false alarm $\alpha_0 \in]0;1[$, the decision threshold $\widetilde{\tau}_{\alpha_0}$ given by:*

$$\widetilde{\tau}_{\alpha_0} = \Phi^{-1}(1 - \alpha_0) \tag{19}$$

where $\Phi^{-1}(\cdot)$ is the Gaussian inverse cumulative distribution, asymptotically warrants that the test $\widetilde{\delta}_2$ (17) is in $\mathcal{K}_{\alpha_0}$.

The main conclusion of Theorem 1 is that the decision threshold $\widetilde{\tau}_{\alpha_0}$ depends neither on the embedding rate R nor the image parameters θ_n and σ_n. Hence, by using the "normalized" log-LR $\widetilde{\Lambda}_2(\mathbf{Z})$, the same threshold permits us to respect a prescribed false-alarm probability α_0 whatever the analyzed image and the embedding rate are.

Equation (18) also implies that to asymptotically calculate the detection power of LR test $\widetilde{\delta}_2$ (17), one only needs to calculate the first moments of $\widetilde{\Lambda}_2(\mathbf{Z})$. The mean expectations used in the log-LR $\widetilde{\Lambda}_2(z_n)$ are given under hypotheses $\mathcal{H}_0$

and $\mathcal{H}_1$ by

$$\mu_0 = \frac{1}{N}\sum_{n=1}^{N}\sum_{k\in\mathcal{Z}} p_{\theta_n}[k]\log\left(\exp\left(\frac{k-\theta_n}{\sigma_n^2}\right)+\exp\left(\frac{\theta_n-k}{\sigma_n^2}\right)\right), \tag{20}$$

$$\mu_2 = \frac{1}{N}\sum_{n=1}^{N}\sum_{k\in\mathcal{Z}} q^R_{\theta_n}[k]\log\left(\exp\left(\frac{k-\theta_n}{\sigma_n^2}\right)+\exp\left(\frac{\theta_n-k}{\sigma_n^2}\right)\right), \tag{21}$$

where the probabilities $p_{\theta_n}[k]$ and $q^R_{\theta_n}[k]$ are respectively defined in (3) and (7). Similarly, the mean variances are by definition given under both hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ by:

$$\sigma_0^2 = \frac{1}{N}\sum_{n=1}^{N}\sum_{k\in\mathcal{Z}} p_{\theta_n}[k]\log\left(\exp\left(\frac{k-\theta_n}{\sigma_n^2}\right)+\exp\left(\frac{\theta_n-k}{\sigma_n^2}\right)\right)^2 - \mu_0^2, \tag{22}$$

$$\sigma_2^2 = \frac{1}{N}\sum_{n=1}^{N}\sum_{k\in\mathcal{Z}} q^R_{\theta_n}[k]\log\left(\exp\left(\frac{k-\theta_n}{\sigma_n^2}\right)+\exp\left(\frac{\theta_n-k}{\sigma_n^2}\right)\right)^2 - \mu_2^2. \tag{23}$$

The expectations μ_0 and μ_2 and the variances σ_0^2 and σ_2^2 as functions of θ_n are respectively drawn in Figures 1a and 1b. These figures highlight the fact that the pixel expectation θ_n can have a significant impact on the LR moments, and later on the detection power, particularly when $\sigma_n < 1$. However, a thorough study of equations (20)–(23) shows that this phenomenon rapidly tends to be negligible when $\sigma_n \gtrsim 1$.
Even thoug, the moments given in (20)–(23) have a rather complicated expression, their numerical calculation is straightforward as long as the parameters θ_n and σ_n are known.

From the asymptotic distribution (18) of the log-LR $\widetilde{\Lambda}_2(\mathbf{Z})$ and the expressions (20)–(23) of its two first moments, the detection power of the LR test $\widetilde{\delta}_2$ (17) is given by the following theorem.

Theorem 2. *For any $\alpha_0 \in]0;1[$, assuming that the parameters $\{\theta_n\}_{n=1}^N$ and $\{\sigma_n\}_{n=1}^N$ are known, the power function $\widetilde{\beta}_{\delta_2}$ associated with the test $\widetilde{\delta}_2$ (17) is asymptotically given, as $N\to\infty$, by:*

$$\widetilde{\beta}_{\delta_2} = 1 - \Phi\left(\frac{\sigma_0}{\sigma_2}\Phi^{-1}(1-\alpha_0) + \frac{\sqrt{N}(\mu_0-\mu_2)}{\sigma_2}\right). \tag{24}$$

Proof. Using the result (18), it asymptotically holds that for any $\widetilde{\tau}_{\alpha_0}\in\mathbb{R}$:

$$\alpha_0(\widetilde{\delta}_2) = \mathbb{P}_0\left[\widetilde{\Lambda}_2(\mathbf{Z}) > \widetilde{\tau}_{\alpha_0}\right] = 1-\Phi\left(\widetilde{\tau}_{\alpha_0}\right).$$

Hence, because Φ is strictly increasing, one has:

$$(1-\alpha_0(\widetilde{\delta}_2)) = \Phi(\widetilde{\tau}_{\alpha_0}) \Leftrightarrow \widetilde{\tau}_{\alpha_0} = \Phi^{-1}\left(1-\alpha_0(\delta_2)\right), \tag{25}$$

which proves Theorem 1.
It also follows from (18) that for any decision threshold $\widetilde{\tau}_{\alpha_0} \in \mathbb{R}$ the power of the test $\widetilde{\delta}_2$ (17) is given by:

$$\widetilde{\beta}_{\delta_2} = \mathbb{P}_1\left[\widetilde{\Lambda}_2(\mathbf{Z}) > \widetilde{\tau}_{\alpha_0}\right] = 1 - \Phi\left(\frac{\sigma_0}{\sigma_2}\left(\widetilde{\tau}_{\alpha_0} - \frac{\sqrt{N}(\mu_2 - \mu_0)}{\sigma_0}\right)\right).$$

By substituting $\widetilde{\tau}_{\alpha_0}$ by the value given in Theorem 1, a short algebra leads to the relation (24). This proves Theorem 2 and concludes the proof.

4.2 General case of $R \in]0;1[$.

The case for which the embedding rate R can take any value in $]0;1]$ is treated in a similar manner as the case $R = 2$. The problem of designing an optimal test has been shown to be particularly difficult in [26]. A thorough design a MP test uniformly with respect to the embedding rate lies outside of the scope of this paper which mainly studies the MP test for $R = 2$ and its practical implementation. Hence, it is proposed to use the test $\widetilde{\delta}_2$ (17) whatever the embedding rate R might be. Once again, the asymptotic distribution (18) is used to solve the decision problem (8).

The alternative hypothesis $\mathcal{H}_R$, that $\mathbf{Z}$ contains a stego-medium with embedding rate $R \in]0;1]$, can be considered as a combination of stego and cover pixels. Hence, the use of the law of total expectation and the law of total variance is relevant to calculate the two first moments of the log-LR $\widetilde{\Lambda}_2(\mathbf{Z})$. Using the moments given in (20)–(23), for the case $R = 2$, a short calculation gives:

$$\mu_R = \frac{R}{2}\mu_2 + \left(1 - \frac{R}{2}\right)\mu_0, \tag{26}$$

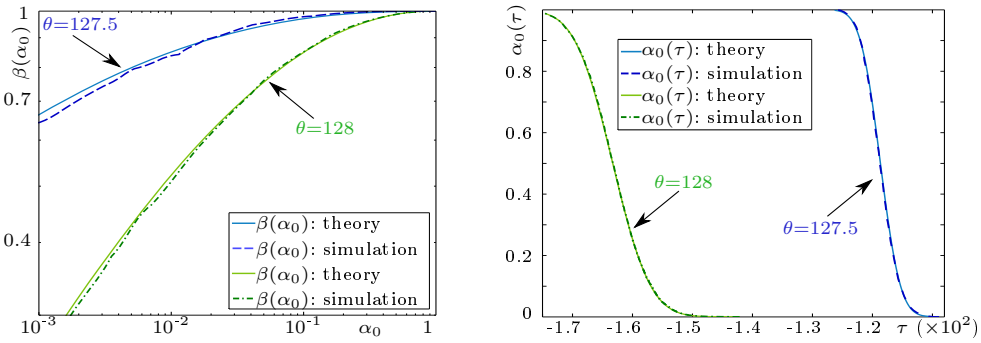
$$\sigma_R^2 = \frac{R}{2}(\sigma_2^2 + \mu_2^2) + \left(1 - \frac{R}{2}\right)(\sigma_0^2 + \mu_0^2) - \left(\frac{R}{2}\mu_2 + \left(1 - \frac{R}{2}\right)\mu_0\right)^2. \tag{27}$$

In other words, by using the test $\widetilde{\delta}_2$ (17) for any $R \in]0;1]$ only the detection power is impacted. Indeed, the null hypothesis does not change, hence, the asymptotic distribution (18) of the LR $\widetilde{\Lambda}_2(\mathbf{Z})$ under $\mathcal{H}_0$ as well as the decision threshold $\widehat{\tau}_{\alpha_0}$ (19) remain the same. This point is highlighted in the following theorem.

Theorem 3. *For any $\alpha_0 \in]0;1[$, assuming that the parameters $\{\theta_n\}_{n=1}^N$ and $\{\sigma_n\}_{n=1}^N$ are known, the power function $\widetilde{\beta}_{\delta_R}$ associated with the test $\widetilde{\delta}_2$ (17) is asymptotically given for any $R \in]0;1]$ by:*

$$\widetilde{\beta}_{\delta_R} = 1 - \Phi\left(\frac{\sigma_0}{\sigma_R}\Phi^{-1}(1 - \alpha_0) + \frac{R\sqrt{N}(\mu_0 - \mu_2)}{\sigma_R}\right). \tag{28}$$

The power functions $\widetilde{\beta}_{\delta_R}$ for $N = 1000$, $R = 0.1$, $\sigma_n = 0.5$ and $\theta_n = \{127.5; 128\}$ are drawn in Figure 2a. Once again, this figure highlights the potentially significant



(a) Detection power as a function of false alarm probability α_0 (ROC curves).

(b) False alarm probability α_0 as a function of threshold value τ_{α_0}.

Fig. 2: Illustration of LRT statistical performance, false-alarm probabilities and detection power, for $N = 1000$ pixels, $R = 0.1$, $\sigma_n = 0.5$ and $\theta = \{127.5; 128\}$. The empirical results were obtained with 5.10^4 realizations.

impact of pixel expectation on the performance of the test $\widetilde{\delta}_2$.
It should be highlighted that the most powerful property of the test $\widetilde{\delta}_2$ is difficult to prove for $R \in]0;1[$, see [9]. However, Figure 3 emphasizes the relevance of the proposed approach, which consists in designing a test for $R = 2$ and extending its application to $R \in]0;1[$. Here, the power function of the proposed test is compared with the power function of the clairvoyant detector, that knows R. The numerical comparison present in Figure 3 shows that the loss of the power is negligible.

Finally, it can be noted that the detection power as given in Theorem 3 complies with the square root law of steganographic capacity [20]. Indeed, from (28), a short algebra immediately permits us to establish that:

$$\lim_{\sqrt{N}/L \to 0} \widetilde{\beta}_{\delta_R} = 1 \quad \text{and} \quad \lim_{\sqrt{N}/L \to \infty} \widetilde{\beta}_{\delta_R} = \alpha_0. \tag{29}$$

5 Practical implementation of proposed LR test.

In a practice, the application of the test $\widetilde{\delta}_2$ (17) is compromised because neither the expectation θ_n nor the variance σ_n^2 of pixels are known: their estimated values, denoted $\widehat{\theta}_n$ and $\widehat{\sigma}_n^2$, respectively, have to be used instead.

However, accurate estimation of the parameters θ_n and σ_n is a difficult problem but necessary to obtain a high detection performance. This problem also occurs in LSB replacement steganalysis. An efficient yet simple way to overcome this problem was introduced in the well-known Weighted Stego-image steganalysis (WS), initially proposed in [14]. The authors propose to locally estimate the parameter θ_n by filtering the inspected image so that $\widehat{\theta}_n$ correspond to the mean of the four surrounding pixels. Similarly, the local variance of the four

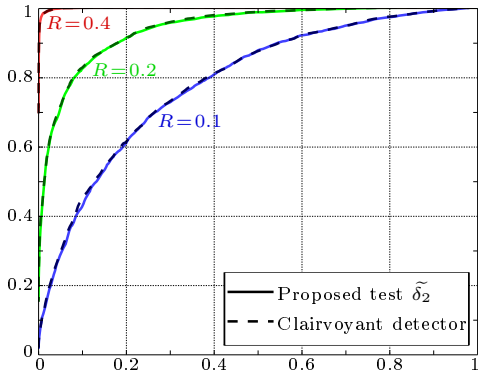


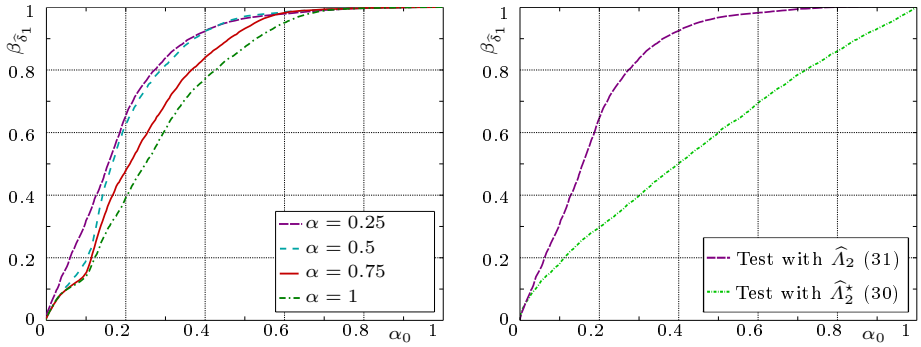
Fig. 3: Numerical comparison between Proposed LR test $\widetilde{\delta_2}$ (17), and the clairvoyant detector which knows the embedding rate $R = 0.1$ ans, thus, uses the LR test design for this rate. Results were obtained from a Monte-Carlo simulation with 5.10^4 realizations using *Lena* image cropped to 128×128 pixels and addition of a Gaussian white noise with $\sigma = 2$.

surrounding pixels is used to estimate σ_n^2. The WS method has been studied thoroughly in [21] and two major improvements have been proposed. First, the authors have empirically enhanced the estimation of pixel expectations by testing different local filters. Second, the author proposed to use moderated weights $w_n = \widehat{\sigma}_n^2 + \alpha$, $\alpha > 0$ instead of the variance estimation $\widehat{\sigma}_n^2$.

In the present paper, it is proposed to use the WS filtering method to estimate the parameters θ_n and σ_n^2. Note that the proposed practical test is not optimal but intends to show the relevance of the proposed approach and feasibility to design a practical efficient test. Following the WS method, the practical implementation of the LR test $\widehat{\delta_2}$ proposed in this paper estimates each θ_n by filtering the inspected image with the kernel:

$$\frac{1}{4}\begin{pmatrix} -1 & 2 & -1 \\ 2 & 0 & 2 \\ -1 & 2 & -1 \end{pmatrix}$$

Contrary to what is suggested in [21], for the case of LSB replacement, our numerical experiments indicate that the detection performance tends to get worse when using the moderated weights instead of the estimated variance. Our interpretation of this phenomenon is as follows. The proposed LR test (17) essentially relies on the increase of pixels' variance due to insertion of hidden information. Hence, the use of moderated weights tends to fundamentally bias the test and deflates the performance results. Figure 4a offers an example of this phenomenon through a comparison of ROC curves obtained using 10 000 images from the BOSSbase database with $R = 1/2$ and $\alpha = \{1/4; 1/2; 3/4; 1\}$.
On the other hand, the direct use of the estimated variance $\widetilde{\sigma}_n^2$ may lead to numerical instability particularly in flat image areas. Hence, it was chosen to add $\alpha = 1/4$ to the estimated variance in our numerical experiments.



(a) ROC obtained with four different weight factor: $\alpha = \{1/4\,;\, 1/2\,;\, 3/4\,;\, 1\}$.

(b) ROC curves obtained with the two statistics $\widehat{\Lambda}_2$ and $\widehat{\Lambda}_2^\star$.

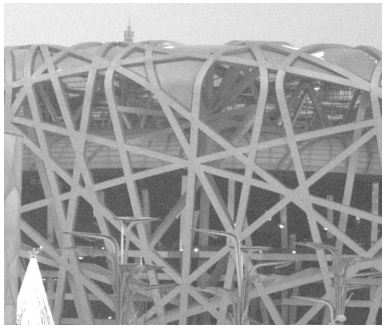
Fig. 4: Impact of weights and calibration on proposed test performance. ROC curves obtained using the images from BOSS database [3] with $R = 0.5$.

By using these estimated values in expression (15) the estimated log-LR $\widetilde{\Lambda}_2(z_n)$, see Equation (15) becomes:

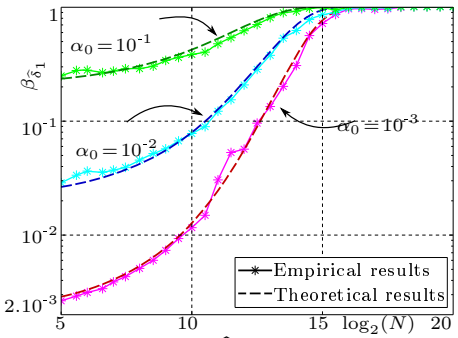
$$\widehat{\Lambda}_2(z_n) = \log\left[\exp\left(\frac{z_n - \widehat{\theta}_n}{(\alpha + \widehat{\sigma}_n)^2}\right) + \exp\left(\frac{\widehat{\theta}_n - z_n}{(\alpha + \widehat{\sigma}_n)^2}\right)\right]. \tag{30}$$

It should be highlighted that some difficult problems still remain open.
First, the normalization of the log-LR, suggested in Equation (17), requires the calculation of the expectation μ_0 and the variance σ_0^2 of the log-LR. Unfortunately, the estimates of the parameters σ_n are, in practice, not accurate enough to perform this normalization efficiently.
Second, possibly the most difficult problem is that the statistical inference between the cover image and the hidden information should be taken into account. For instance it was proposed in [26] to remove the LSB plane in order to remove any potential stego-noise. For LSB matching this is not possible. Therefore, the impact of hidden information on estimators $\widehat{\theta}_n$ and $\widehat{\sigma}_n$ should be studied. Since the proposed test relies mainly on the slight increase of pixels' variance due to data hiding, the embedding changes may have an important effect on the estimates $\widehat{\sigma}_n$ and on the proposed test.

As explained above, proper normalization of the proposed test is critical in practice. Even though the proposed LR is very sensitive to hidden information, if its expectation can not be set to a fixed value under $\mathcal{H}_0$, the between-image-error described in [2] may negatively impact the test accuracy. Numerical simulations show that the expectation of the LR $\widehat{\Lambda}_2(z_n)$ can be roughly approximated by $-\log(2) - \frac{1}{4\widehat{\sigma}_n^2}$.



(a) Digital image used for the Monte-Carlo simulations



(b) Power of the test $\widehat{\delta}_2$ (31) as a function of pixel number for different false-alarm probabilities: theory and simulation.

Fig. 5: Numerical verification of theoretical results through Monte-Carlo simulation based on natural image shown in Figure 5a.

Therefore, the practical test proposed in the present paper is given as:

$$\widehat{\delta}_2 = \begin{cases} \mathcal{H}_0 & \text{if} \quad \widehat{\Lambda}_2^\star(\mathbf{Z}) \leq \widehat{\tau}_{\alpha_0}, \\ \mathcal{H}_1 & \text{if} \quad \widehat{\Lambda}_2^\star(\mathbf{Z}) > \widehat{\tau}_{\alpha_0}, \end{cases} \tag{31}$$

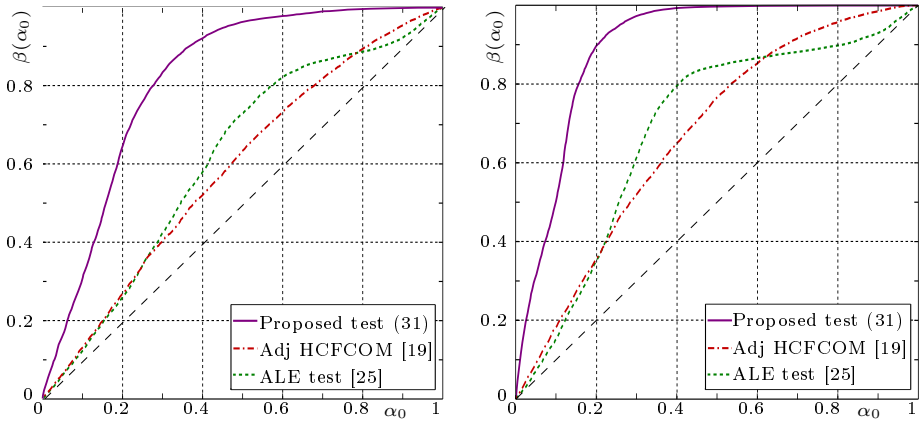
$$\text{with} \quad \widehat{\Lambda}_2^\star(\mathbf{Z}) = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} \widehat{\Lambda}_2(z_n) - \log(2) - \frac{1}{4(\alpha + \widehat{\sigma}_n)^2}. \tag{32}$$

One can note that, contrary to the LR statistically studied throughout Sections 4.1–4.2, the proposed decision statistic is not normalized. Indeed the variance of $\widehat{\Lambda}_2(z_n)$ is not taken into account in Equation 31. This is because the estimation of pixels' variance is particularly difficult and the method used in this paper is not accurate enough. In fact, normalization can even lower the detection performance. The most notable thing about the test (31) is that the expectation of the decision statistics $\widehat{\Lambda}_2^\star(\mathbf{Z})$ is always 0 under hypothesis $\mathcal{H}_0$. Figure 4b shows an example of the detection power obtained with the two tests based on the statistics (30) and (31).

6 Numerical Simulations.

6.1 Theoretical results on simulated data.

Figure 5 presents a numerical verification of Theorem 3. The image shown in Figure 5a has been analyzed 5.10^4 times. Each run was preceded by the addition of a zero-mean Gaussian noise whose standard deviation was $\sigma = 1$. The embedded hidden information was drawn from a binomial distribution $\mathcal{B}(1, 1/2)$ with an embedding rate $R = 1$. The empirical power of the test $\widehat{\delta}_2$ is compared with



(a) ROC curves for $R = 0.5$. (b) ROC curves for $R = 1$.

Fig. 6: Numerical comparisons of detectors performance using BOSS database [3].

the theoretical result given by Theorem 3 for three different false-alarm probabilities: $\alpha_0 = \{10^{-1}; 10^{-2}; 10^{-3}\}$. Observe that the obtained detection power almost perfectly corresponds to the theoretical results.
Note that it is crucial to use the same image for this Monte-Carlo simulation because the detection power of the proposed test depends on image parameters, namely on θ_n and particularly on σ_n^2. Hence, for a different image, the detection power may differ significantly as explained in Section 4. Moreover, the use of the same image artificially permits us to overcome the difficult problem of normalizing the log-LR and, thus, the effects of the between-image-error described in [2].

6.2 Comparison with the state of the art on real images.

Matlab source code of proposed test, as detailed in Equation (31), is available on the Internet at : `http://remi.cogranne.pagesperso-orange.fr/`.

One of the main motivations for this paper was to show that the hypothesis testing theory can be applied in practice to design an efficient LSB matching detector. This fact can only be shown by a numerical comparison with state-of-the-art detectors on large image databases. The potential competitors for LSB matching detection are not as numerous as for LSB replacement. As briefly described in the introduction, the operational context selected in this paper eliminates all prior-art detectors based on machine learning. Almost every other detector found in the literature is based on the image histogram. For the present comparison, two histogram-based detectors, namely ALE [25] and the adjacency HCF COM [19] detector, were used due to their high detection performance.
Figure 6 shows the results obtained with 10 000 images from BOSSbase contest database [3]. Each hidden bit was drawn from a binomial distribution $\mathcal{B}(1, 1/2)$. The embedding rate was $R = 0.5$ in Figure 6a and $R = 1$ in Figure 6b. Both

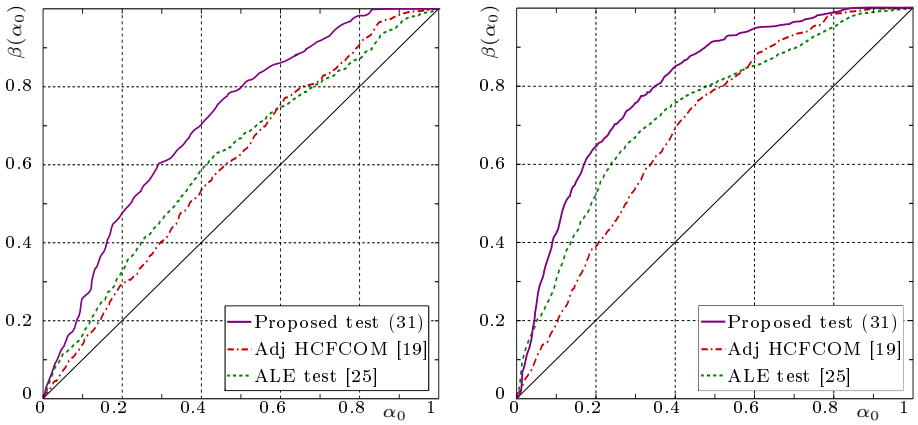
figures show that the proposed test achieves a better detection power for any prescribed false-alarm probability.

Similarly, Figure 7 shows the results obtained with the 1488 raw images from the 'Dresden Image Database' [16]. Prior to our experiments, each image was converted to an unprocessed TIFF format (using dcraw) and only the red color channel was used. The embedding rate was $R = 0.25$ in Figure 7a and $R = 0.5$ in Figure 7b. The results presented in Figures 7a and 7b confirm that the proposed test has a better detection power for any prescribed false-alarm probability. Moreover by changing the embedding rate, the combined results of Figures 6 and 7 show that the proposed test also performs better than prior art for any R.

Note that, surprisingly, the detection power of the proposed test is slightly higher for the BOSSbase database than for the Dresden database for $R = 0.5$, see Figure 6a and 7b, respectively, whereas the Dresden database images are bigger. This phenomenon can be explained by the fact that the Dresden database images are RAW images that have not being further processed. In contrast, BOSSbase images have been downsampled, which may introduce correlations between neighboring pixels that implicitly make the filtering estimator more efficient.

7 Conclusion and future works.

The first step to fill the gap between hypothesis testing theory and steganalysis was recently proposed in [12, 7, 26]. This paper extends this first step to the case of LSB matching. By casting the problem of LSB matching steganalysis in the framework of hypothesis testing theory, the most powerful likelihood ratio test is designed. Then, a thorough statistical study permits analytical calculations



(a) ROC curves for $R = 0.25$.

(b) ROC curves for $R = 0.5$.

Fig. 7: Comparisons of detectors performance using Dresden database [16].

of its performance in terms of the false-alarm probability and detection power. To apply this test in practice, unknown image parameters have to be estimated. Based on a simple estimation of these unknown parameters, a practical test is proposed.
The relevance of the proposed approach is emphasized through numerical experiments. Compared to two leading histogram-based detectors, the proposed practical test achieves a better detection power.

However, the practical test presented in this paper relies on a simple yet efficient filtered version of inspected media to estimate pixel expectations and variances. In our future work, a more efficient model should be used to increase the detection power. Lastly, a thorough statistical study of the impact of this estimation on detection performance is desirable to complete the present work.

8 Acknowledgments.

The authors would like to thank Jessica Fridrich for her contributions and stimulating discussions.

References

1. Böhme, R.: Advanced Statistical Steganalysis. Springer Publishing Company, Incorporated, 1st edition. (2010)
2. Böhme, R., Ker, A.D.: A two-factor error model for quantitative steganalysis. In: Security, Steganography, and Watermarking of Multimedia Contents VIII Proc. of the SPIE, vol. 6072 (2006)
3. BOSS contest: Break Our Steganographic System (2010), `http://www.agents.cz/boss/`
4. Cai, K., Li, X., Zeng, T., Yang, B., Lu, X.: Reliable histogram features for detecting LSB matching. In: Image Processing (ICIP), 2010 17th IEEE International Conference on. pp. 1761 –1764 (Sept. 2010)
5. Cancelli, G., Doerr, G., Barni, M., Cox, I.: A comparative study of ± 1 steganalyzers. In: Multimedia Signal Processing, 2008 IEEE 10th Workshop on. pp. 791 –796 (Oct. 2008)
6. Cancelli, G., Doerr, G., Cox, I., Barni, M.: Detection of ± 1 LSB steganography based on the amplitude of histogram local extrema. In: Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on. pp. 1288 –1291 (Oct. 2008)
7. Cogranne, R., Zitzmann, C., Fillatre, L., Nikiforov, I., Retraint, F., Cornu, P.: A cover image model for reliable steganalysis. In: Information Hiding. pp. 178 – 192. LNCS vol.6958, Springer (2011)
8. Cogranne, R., Zitzmann, C., Fillatre, L., Nikiforov, I., Retraint, F., Cornu, P.: Reliable detection of hidden information based on a non-linear local model. In: Statistical Signal Processing, Proc. of IEEE Workshop on. pp. 493 – 496 (2011)
9. Cogranne, R., Zitzmann, C., Fillatre, L., Retraint, F., Nikiforov, I., Cornu, P.: Statistical decision by using quantized observations. In: IEEE International Symposium on Information Theory. pp. 1135 – 1139 (2011)

10. Cogranne, R., Zitzmann, C., Nikiforov, I., Retraint, F., Fillatre, L., Cornu, P.: Statistical Detection of LSB Matching in the Presence of Nuisance Parameters. In: *accepted for publication in* Statistical Signal Processing, Proc. of IEEE Workshop on (2012)
11. Cox, I., Miller, M., Bloom, J., Fridrich, J., Kalker, T.: Digital Watermarking and Steganography. Morgan Kaufmann, 2nd edition. (2007)
12. Dabeer, O., Sullivan, K., Madhow, U., Chandrasekaran, S., Manjunath, B.: Detection of hiding in the least significant bit. Signal Processing, IEEE Transactions on 52(10), 3046 – 3058 (Oct. 2004).
13. Foi, A., Trimeche, M., Katkovnik, V., Egiazarian, K.: Practical Poissonian-Gaussian noise modeling and fitting for single-image raw-data. Image Processing, IEEE Transactions on 17(10), 1737–1754 (Oct. 2008)
14. Fridrich, J., Goljan, M.: 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, vol. 5306 (2004)
15. Fridrich, J.: Steganography in Digital Media: Principles, Algorithms, and Applications. Cambridge University Press, 1st edition (2009)
16. Gloe, T., Böhme, R.: The 'Dresden Image Database' for benchmarking digital image forensics. In: Proceedings of the 25th Symposium On Applied Computing (ACM SAC 2010). vol. 2, pp. 1585–1591 (2010)
17. Goljan, M., Fridrich, J., Holotyak, T.: New blind steganalysis and its implications. In: Security, Steganography, and Watermarking of Multimedia Contents VIII Proc. of the SPIE, vol. 6072 (2006)
18. Harmsen, J., Pearlman, W.: Higher-order statistical steganalysis of palette images. In: Security, Steganography, and Watermarking of Multimedia Contents V, Proc. of the SPIE, vol. 5020 (2005)
19. Ker, A.: Steganalysis of LSB matching in grayscale images. Signal Processing Letters, IEEE 12(6), 441 – 444 (June 2005)
20. Ker, A.D.: A capacity result for batch steganography. Signal Processing Letters 14(8), 525–528 (2007)
21. Ker, A.D., Böhme, R.: Revisiting weighted stego-image steganalysis. In: Security, Forensics, Steganography, and Watermarking of Multimedia Contents X, Proc. of the SPIE, vol. 6819 (2008)
22. Lehman, E., Romano, J.: Testing Statistical Hypotheses, Second Edition. Springer, 3rd edition. (2005)
23. Lyu, S., Farid, H.: Steganalysis using higher-order image statistics. Information Forensics and Security, IEEE Transactions on 1(1), 111 – 119 (March 2006).
24. Scott, C.: Performance measures for Neyman-Pearson classification. IEEE Trans. Inform. Theory 53(8), 2852–2863 (2007)
25. Zhang, J., Cox, I., Doerr, G.: Steganalysis for LSB matching in images with high-frequency noise. In: Multimedia Signal Processing, 2007. MMSP 2007. IEEE 9th Workshop on. pp. 385 –388 (Oct. 2007).
26. Zitzmann, C., Cogranne, R., Retraint, F., Nikiforov, I., Fillatre, L., Cornu, P.: Statistical decision methods in hidden information detection. In: Information Hiding. pp. 163 – 177. LNCS vol.6958, Springer (2011).