RESEARCH

Steganalysis of JSteg Algorithm Using Hypothesis Testing Theory

Tong Qiao*, Florent Retraint, Rémi Cogranne and Cathel Zitzmann
[1]

Abstract

This paper investigates the statistical detection of JSteg steganography. The approach is based on a statistical model of Discrete Cosine Transformation (DCT) coefficients challenging the usual assumption that among a subband all the coefficients are independent and identically distributed (i. i. d.). The hidden information detection problem is cast in the framework of hypothesis testing theory. In an ideal context where all model parameters are perfectly known, the Likelihood Ratio Test (LRT) is presented and its performances are theoretically established. The statistical performance of LRT serves as an upper bound for the detection power. For a practical use where the distribution parameters are unknown, by exploring a DCT channel selection, a detector based on estimation of those parameters is designed. The loss of power of the proposed detector, compared with the optimal LRT is small, shows the relevance of the proposed approach.

Keywords: Hypothesis testing theory; JSteg steganalysis; DCT distribution model; hidden information detection

1 Introduction and contributions

Steganography and steganalysis have received more and more focus in the past two decades since the research in this field concerns law enforcement and national strategic defense. Steganography is the art and science of hiding secret messages in the cover media. On the opposite, steganalysis is about the detection of hidden secret information embedded in the cover media, also called stego media. If a steganalysis algorithm detects the inspected media as the stego one, even without knowing any extra information about the secret message, steganographic approach fails.

1.1 State of the art

In today's digital world, there exists many steganographic tools available on the Internet. Due to the fact that some are readily available and very simple to use, it is necessary to design the most reliable steganalysis methodology to fight back steganography. In general, due to its simplicity most of steganographic schemes insert the secret message into the Least Significant Bit (LSB) plane of cover media, including two kinds of steganography: LSB replacement and LSB matching. The former algorithm aims at replacing LSB plane in the spatial domain or frequency domain of the cover media by 0 or 1. The latter algorithm, also known as ± 1 embedding (see [1, 2, 3]), randomly increment or decrement pixel or DCT coefficient value to match the secret bit to be embedded when necessary. Since LSB replamcement is easier to implement it remains more popular and, hence, as of December 2011, WetStone declared that about 70 percent of the available steganographic softwares are based on the LSB replacement algorithm [4]. Therefore the research on LSB replacement steganalysis remains an active topic.

Although the LSB replacement steganalysis method (see [5, 6, 7, 8, 9, 10]) has been studied for many years, it can be noted that most of the prior-art detectors are designed to detect data hidden in the spatial domain. In addition, for only a few detectors the statistical properties have been studied and established, referred to as the optimal detectors. As detailed in [11], a wide range of problems, theoretical as well as practical, remain uncovered and some prevent the moving of "steganography and steganalysis from the laboratory into the real world". This is especially the case in the field of Optimal Detection, see [11, Sec. 3.1], in which this paper lies. Roughly speaking, the goal of optimal detection in steganalysis, is to exploit an accurate statistical model of cover source, usually digital images, to design a statistical test whose properties

*Correspondence: tong.qiao@utt.fr
ICD - LM2S - Université de Technologie de Troyes (UTT) - UMR STMR CNRS, 12 rue Marie Curie, 10004 Troyes, FRANCE
Full list of author information is available at the end of the article

can be established; typically, in order to guarantee a False Alarm Rate (FAR) and to calculate the optimal detection performance one can expect from the most powerful detector.

In 2004, the Weighted Stego-image (WS) method [12] and the test proposed in [13] for LSB replacement steganalysis changed the situation opening the way to optimal detectors. Driven by these pioneer works, the enhanced WS algorithm proposed in [14] improved the detection rate by enhancing pixels predictor, adjusting weighting factor and introducing the concept of bias correction. Nevertheless, the drawback of original WS method is that it can only be applied in the spatial domain. Due to the prevalence of images compressed in the Joint Photographic Experts Group (JPEG) format, how to deal with this kind of images becomes mandatory. Inspired by the prior studies [12, 14], the WS steganalyser for JPEG covers was proposed in [15]. However, the WS steganalyser does not allow to get high detection performance for a low FAR, see [16], and its statistical properties remain unknown, which prevents the guarantee of a prescribed FAR. In practical forensic cases, since a large database of images needs to be processed, the getting of a very low FAR is crucial.

1.2 Contributions of the paper

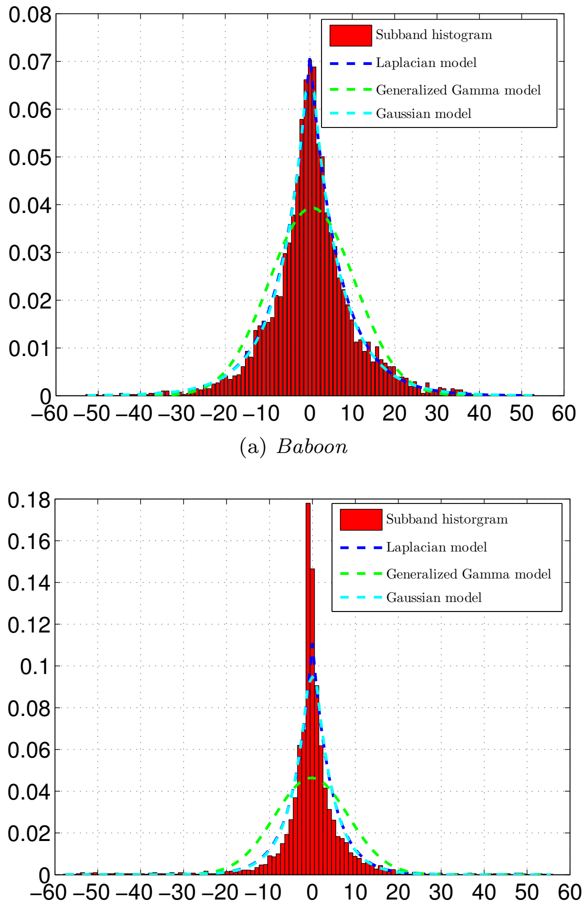
For the detection of data hidden within the Discrete Cosine Transformation (DCT) coefficients of JPEG images, the application of hypothesis testing theory for designing optimal detectors, that are efficient in practice, is facing the problem of accurately modelling statistical distribution of DCT coefficients. It can be noted that several models have been proposed in the literature to model statistically the DCT coefficients. Among those models, the Laplacian distribution is probably the most widely used due to its simplicity and its fairly good accuracy [17]. More accurate models such as the Generalized Gaussian [18] and, more recently, the Generalized Gamma model [19] have been shown to provide much more accuracy at the cost of higher complexity. Some of those models have been exploited in the field of steganalysis, see [20, 21] for instance. In the framework of optimal detection, a first attempt has been made to design a statistical test modelling the DCT coefficient with the quantized Laplacian distribution, see [22].

It should be noted that other approaches have been proposed for the detection of data hidden within DCT coefficients of JPEG images, to cite a few, the structural detection [23], the category attack [24], the WS detector [15], and universal or blind detectors [25, 26]. However, establishing the statistical properties of those detectors remains a difficult work which has not been studied yet. In addition, most accurate detectors based on statistical learning are sensitive to the so-called cover source-mismatch [27]: the training phase must be performed with caution.

In this context, the detector proposed in [22] is an interesting alternative; however it is based on the assumption that DCT coefficients are independent and identically distributed (i. i. d.) within a subband and have a zero expectation which might be inaccurate and hence make the detection performance poor in practice. In practice, this model is not independent of the image content, which performs well only in the case of high-texture image (See Figure 1a), but hardly holds true in the case of low-texture image (See Figure 1b). On the opposite, this paper proposes a statistical model assuming that each DCT coefficient has a different expectation and variance. The use of this model, together with hypothesis theory, allows us to design the most powerful Likelihood Ratio Test (LRT) when the distribution parameters (expectation and variance) are known. Then in the practical case of not knowing those parameters, estimations have to be used instead; this leads to the design of the proposed detector with estimated parameters. By taking into account those distribution parameters as nuisance parameters and using an accurate estimation, it is shown that the loss of power compared with the optimal detector is small.

Therefore, the contributions of this paper are as follows:

1. First, a novel model of DCT coefficients is proposed ; its major originality is that this model does not assume that all the coefficients of the same subband are i. i. d. .
2. Second, assuming that all the parameters are known, this statistical model of DCT coefficients is used to design the optimal test to detect data hidden within JPEG images with JSteg algorithm. This statistical test takes into account distribution parameters of each DCT coefficient as nuisance parameters.
3. Further, assuming that all the parameters are unknown, a simple approach is proposed to estimate the expectation (or location) parameter of each coefficient by using linear properties of DCT as well as estimation of pixels expectation in spatial domain; the variance (or scale) parameter is also estimated locally.
4. The designed detector is improved by exploring a DCT channel selection, which has been proposed very recently [28, 29], that selects only a sub-set of pixels or DCT coefficients in which embedding is most likely and hence detection easier.
5. Numerical results show the sharpness of the theoretically established results and the good per-



(a) *Baboon*

(b) *Sky*

Figure 1: **Illustration of the quantized DCT coefficient subband (2,1), (a) a high-texture image: *Baboon*, (b) a low-texture image: *Sky*.**

formance of the proposed statistical test. A comparison with the statistical test based on the Laplacian distribution and on the assumption of i. i. d. coefficient, see [22], shows the relevance of the proposed methodology. In addition, compared with prior-art WS detector [15], experimental results show the efficiency of the proposed detector.

1.3 Organization of the paper

This paper is organised as follows. Section 2 formalises the statistical problem of detection of information hidden within DCT coefficients of JPEG images. Then, Section 3 presents the optimal Likelihood Ratio Test (LRT) for detecting the JSteg algorithm based on the Laplacian distribution model. Section 4 presents the proposed approach for estimating the nuisance parameters in practice, and compares our proposed detector with WS detector [15] theoretically. Finally, Section 5 presents numerical results of the proposed steganalyser on simulated and real images and Section 6 concludes this paper. This paper is an extended version of [30] that also includes the findings of [31] on channel selection [28, 29].

2 Problem Statement

In this paper, a grayscale digital image is represented, in the spatial domain, by a single matrix $\mathbf{Z} = \{z_{i,j}\}, i \in \{1, \ldots, I\}, j \in \{1, \ldots, J\}$. The present work can be extended to colour image by analysing each colour channel separately. Most of digital images are stored using the JPEG compression standard. This standard exploits the linear Discrete Cosine Transform (DCT), over blocks of 8×8 pixels to represent an image in the so-called DCT domain. In the present paper, we avoid the description of the imaging pipeline of a digital still camera; the reader can refer to [32] for a description of the whole imaging pipeline and to [33] for a detailed description of the JPEG compression standard.

Let us denote DCT coefficients by the matrix $\mathbf{V} = \{v_{i,j}\}$. An alternative representation of those coefficients is usually adopted by gathering the DCT coefficients that corresponds to the same frequency subband. In this paper, this alternative representation is denoted by the matrix $\mathbf{U} = \{u_{k,l}\}, k \in \{1, \ldots, K\}, l \in \{1, \ldots, 64\}$ with $K \approx I \times J/64$[2].

The coefficients from the first subband $u_{k,1}$, often referred to as DC coefficients, represent the mean of pixels value over k-th block of 8×8 pixels. The modification of those coefficients may be obvious and creates artifacts that can be detected easily, hence, they are usually not used for data hiding. Similarly, the JSteg algorithm does not use the coefficients from the other subbands, referred to as AC coefficients, if they equal 0 or 1. In fact, it is known that using the coefficients equal to 0 or 1 modifies significantly statistical properties of AC coefficients; this creates a flaw that can be detected.

The JSteg algorithm embeds data within DCT coefficients of JPEG images using the well-known LSB (Least Significant Bit) replacement method, see details in [34]. In brief, this method consists in substituting the LSB of each DCT coefficient by a bit of the message it is aimed to hide. The number of bit hidden per coefficient, usually referred to as the payload, is denoted $R \in (0, 1]$. Since the JSteg algorithm does not use each DCT coefficient, the payload will in fact be measured in this paper as the number of bits hidden

[2] In this paper we assume, without loss of generality, that both width and height of inspected image are multiples of 8.

per *usable* coefficients (that is the number of bits divided by the number of AC coefficients that differ from 0 and 1).

Let us assume that the DCT coefficients are independent and that they all follow the same probability distribution, denoted $\mathcal{P}_\theta$, parametrised by the parameter θ which may change among the coefficients. Since the DCT coefficients can only take value into a discrete set, the distribution $\mathcal{P}_\theta$ may be represented by its probability mass function (pmf) denoted $P_\theta = \{p_\theta[u]\}$; for simplicity[3], it is assumed in this paper that $u \in \mathbb{Z}$. Let us denote $\mathcal{Q}_\theta^R$ the probability distribution of *usable* DCT coefficients from the stego-image, after embedding a message with payload R. A short calculation shows that, see [13, 12, 8], the stego-image distribution may be represented with following the pmf $Q_\theta^R = \{q_\theta^R[u]\}_{u\in\mathbb{Z}}$ where

$$q_\theta^R[u] = (1 - {}^R\!/\!_2)p_\theta[u] + {}^R\!/\!_2 p_\theta[\bar{u}], \tag{1}$$

and $\bar{u} = u+(-1)^u$ represents the integer u with flipped LSB. For the sake of clarity, let us denote $\theta_{k,l}$ the distribution parameter of k-th DCT coefficient from l-th subband and let $\boldsymbol{\theta} = \{\theta_{k,l}\}, k \in \{1,\dots,K\}, l \in \{2,\dots,64\}$ represents the distribution parameter of all the AC coefficients.

When inspecting a given JPEG image, more precisely its DCT coefficients matrix $\mathbf{U}$, in order to detect data hidden with the JSteg algorithm, the problem consists in choosing between the two following hypotheses $\mathcal{H}_0$: "*the coefficients $u_{k,l}$ follow the distribution $\mathcal{P}_{\theta_{k,l}}$*" and $\mathcal{H}_1$: "*the coefficients $u_{k,l}$ follow the distribution $\mathcal{Q}_{\theta_{k,l}}^R$*" which can be written formally as:

$$\begin{cases} \mathcal{H}_0 : \left\{ u_{k,l} \sim \mathcal{P}_{\theta_{k,l}}, \forall k \in \{1,\dots,K\}, \forall l \in \{2,\dots,64\} \right\}, \\ \mathcal{H}_1 : \left\{ u_{k,l} \sim \mathcal{Q}_{\theta_{k,l}}^R, \forall k \in \{1,\dots,K\}, \forall l \in \{2,\dots,64\} \right\}. \end{cases} \tag{2}$$

A statistical test is a mapping $\delta : \mathbb{Z}^{I\cdot J} \mapsto \{\mathcal{H}_0, \mathcal{H}_1\}$ such that hypothesis $\mathcal{H}_i$ is accepted if $\delta(\mathbf{U}) = \mathcal{H}_i$ (see [35] for details on hypothesis testing). As previously explained, this paper focuses on the Neyman-Pearson bi-criteria approach: maximising the correct detection probability for a given false-alarm probability α_0. Let:

$$\mathcal{K}_{\alpha_0} = \left\{ \delta : \sup_{\boldsymbol{\theta}} \mathbb{P}_{\mathcal{H}_0}[\delta(\mathbf{U}) = \mathcal{H}_1] \le \alpha_0 \right\}, \tag{3}$$

be the class of tests with a false alarm probability upper-bounded by α_0. Here $\mathbb{P}_{\mathcal{H}_i}(A)$ stands for the probability of event A under hypothesis $\mathcal{H}_i, i = \{0,1\}$, and the supremum over $\boldsymbol{\theta}$ has to be understood as whatever the distribution parameters might be, in order to ensure that the false alarm probability α_0 can not be exceed.
Among all the tests in $\mathcal{K}_{\alpha_0}$, it is aimed at finding a test δ which maximises the power function, defined by the correct detection probability:

$$\beta_\delta = \mathbb{P}_{\mathcal{H}_1}[\delta(\mathbf{U}) = \mathcal{H}_1], \tag{4}$$

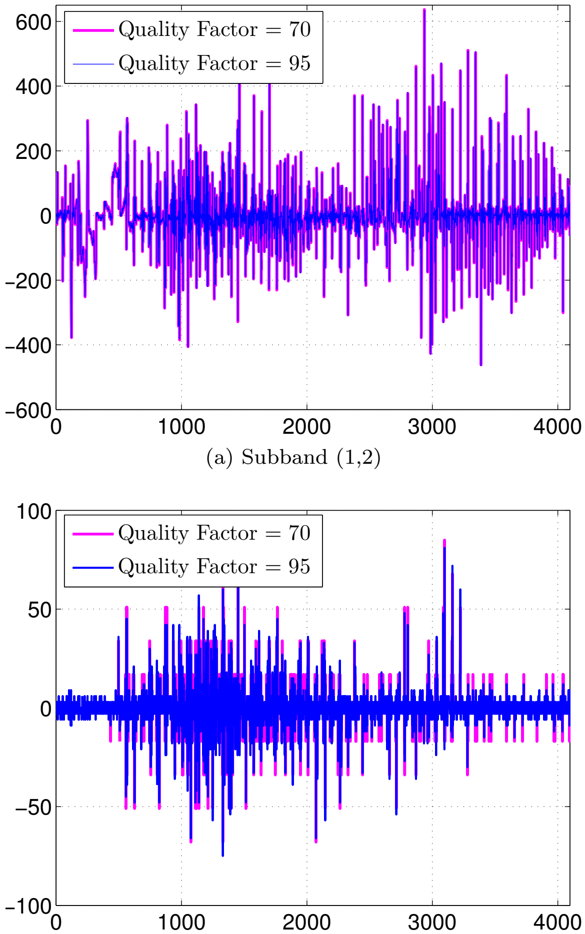
which is equivalent to minimize the missed detection probability $\alpha_1(\delta) = \mathbb{P}_{\mathcal{H}_1}[\delta(\mathbf{U}) = \mathcal{H}_0] = 1 - \beta_\delta$.

In order to design a practical *optimal detector*, as referred in [11], for steganalysis in spatial domain, the main difficulty is to estimate the distribution parameters (expectation and variance of each pixel). On the opposite, in the case of DCT coefficients, the application of hypothesis testing theory to design an optimal detector has previously being attempted with the assumption that the distribution parameter remains the same for all the coefficients from a same subband. With this assumption, the estimation of the distribution parameters is not an issue because thousands of DCT coefficients are available. However which distribution model to choose remains an open problem.

The hypothesis testing theory has been applied for the steganalysis of JSteg algorithm in [22] using a Laplacian distribution model and using the assumption that DCT coefficients of each subband are i. i. d.. However, this pioneer work does not allow the designing of an efficient test because a very important loss of performance has been observed when comparing results on real images and theoretically established ones. Such a result can be explained by the two following reasons: 1) the Laplacian model might be not accurate enough to detect steganagraphy and 2) the assumption that the DCT coefficients of each frequency subband are i. i. d. may be wrong. Recently, it has been shown that the use of Generalised Gamma model or even more accurate model [36, 37] allows the designing of a test with very good detection performance. On the opposite, in this paper, it is proposed to challenge the assumption that all the DCT coefficients of a subband are i. i. d..

A typical example is given by Figure 2 and 3. Figure 2a (resp. Figure 2b) represents the DCT coefficients of the subband (1,2) (resp. subband (4,4)) extracted from the image *lena*. Observing those two graphs, it is obvious that the assumption of all those coefficients being i. i. d. is doubtful. However, if it is assumed that each coefficient has a different expectation, one can estimate this expected value and compute the "residual noise", that is the difference between the

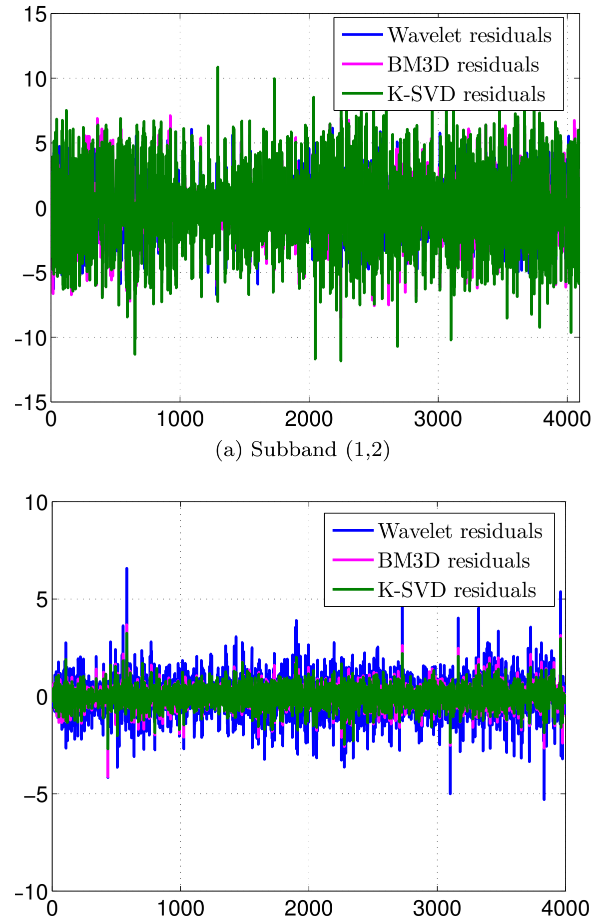
[3] In practice, DCT coefficients belong to set $[-1024, \dots, 1023]$, see [22].



(a) Subband (1,2)

(b) Subband (4,4)

Figure 2: **Illustrative examples of the value of DCT coefficients of two subbands from *lena* image. Those examples show that the assumption that DCT coefficients are i. i. d. within a subband hardly holds true in practice.**



(a) Subband (1,2)

(b) Subband (4,4)

Figure 3: **Illustrative examples of DCT coefficients of residual noise, obtained by denoising image. The same two DCT subbands, as in Figure 2 are extracted the residual noise of *lena* image. On those examples, the assumption of i.i.d. distribution seems to be more realistic.**

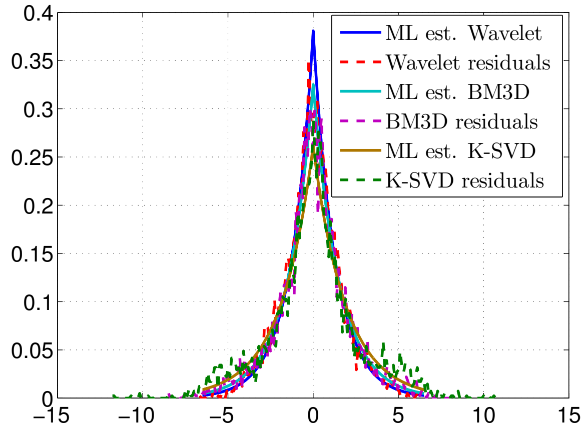
observation and the computed expectation. Such results are shown in Figure 3, with two different models for estimating the expectation of DCT coefficients of the same two subbands from *lena*. Obviously, residual noises look much more i. i. d. than the original DCT coefficients.

In the following section, we detail the statistical test that takes into account both the expectation and the variance as nuisance parameters and we study the optimal detection when those parameters are known. A discussion on nuisance parameters is also provided in Section 4.

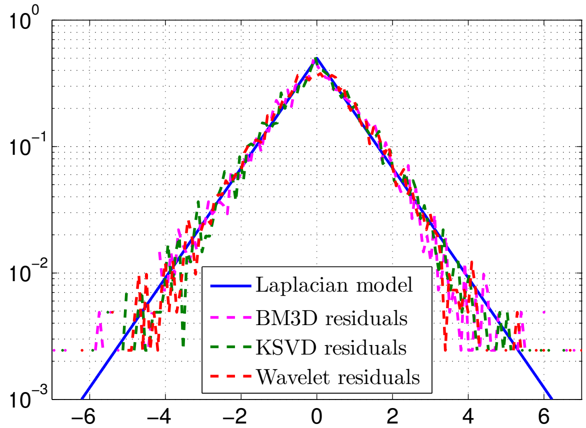
3 LRT for two simple hypothesis

3.1 Optimal detection framework

When the payload R and the distribution parameters $\boldsymbol{\theta} = \{\theta_{k,l}\}, k \in \{1, \ldots, K\}, l \in \{2, \ldots, 64\}$ are known, problem (2) is reduced to a statistical test between two simple hypotheses. In such a case, the Neyman-Pearson Lemma [35, theorem 3.2.1] states that the most powerful test in the class $\mathcal{K}_{\alpha_0}$ (3) is the LRT defined, on the assumption that DCT coefficients are



(a) Subband (1,2)



(b) Subband (4,4)

Figure 4: **Statistical distribution of the DCT coefficients of the residual noise plotted in Figure 4a. For comparison, the Laplacian PDF with parameters estimated by the MLE are also shown. Note that for a meaning comparison, Figure 4b shows the results after normalization by the estimated scale parameter $\widehat{b}_k$.**

independent, as:

$$\delta^{\rm lr}(\mathbf{U}) = \begin{cases} \mathcal{H}_0 \text{ if } \Lambda^{\rm lr}(\mathbf{U}) = \displaystyle\sum_{k=1}^{K}\sum_{l=2}^{64}\Lambda^{\rm lr}(u_{k,l}) < \tau^{\rm lr}, \\ \mathcal{H}_1 \text{ if } \Lambda^{\rm lr}(\mathbf{U}) = \displaystyle\sum_{k=1}^{K}\sum_{l=2}^{64}\Lambda^{\rm lr}(u_{k,l}) \geq \tau^{\rm lr}, \end{cases} \tag{5}$$

where the decision threshold $\tau^{\rm lr}$ is the solution of the equation $\mathbb{P}_{\mathcal{H}_0}\left[\Lambda^{\rm lr}(\mathbf{U}) \geq \tau^{\rm lr}\right] = \alpha_0$, to ensure that the false alarm probability of the LRT equals α_0, and the log Likelihood Ratio (LR) for one observation is given, by definition, by:

$$\Lambda^{\rm lr}(u_{k,l}) = \log\left(\frac{q^R_{\theta_{k,l}}[u_{k,l}]}{p_{\theta_{k,l}}[u_{k,l}]}\right). \tag{6}$$

In practice, when the rate R is not known one can try to design a test which is locally optimal around a given payload rate, named Locally Asymptotically Uniformly Most Powerful (LAUMP) test, as proposed in [6, 8] but this lies outside the scope of this paper.

From the definition of $p_{\theta_{k,l}}[u_{k,l}]$ and $q^R_{\theta_{k,l}}[u_{k,l}]$ (1), it is easy to write the LR (6) as:

$$\Lambda^{\rm lr}(u_{k,l}) = \log\left(1 - \frac{R}{2} + \frac{R}{2}\frac{p_{\theta_{k,l}}[\bar{u}_{k,l}]}{p_{\theta_{k,l}}[u_{k,l}]}\right), \tag{7}$$

where, as previously defined, $\bar{u}_{k,l} = u_{k,l} + (-1)^{u_{k,l}}$ represents the DCT coefficient $u_{k,l}$ with flipped LSB.

3.2 Statistical performance of LRT

Accepting, for a moment, that one is in this most favourable scenario, in which all the parameters are perfectly known, we can deduce some interesting results. Due to the fact that observations are considered to be independent, the LR $\Lambda^{\rm lr}(\mathbf{U})$ is the sum of random variables and some asymptotic theorems allow to establish its distribution when the number of coefficients becomes "sufficiently large". This asymptotic approach is usually verified in the case of digital images due to the very large number of pixels or DCT coefficients.

Let us denote $E_{\mathcal{H}_i}(\theta_{k,l})$ and $V_{\mathcal{H}_i}(\theta_{k,l})$ the expectation and the variance of the LR $\Lambda^{\rm lr}(u_{k,l})$ under hypothesis $\mathcal{H}_i, i = \{0,1\}$. Those quantity obviously depend on the parameterized distribution $\mathcal{P}_{\theta_{k,l}}$. The Lindeberg's central limit theorem (CLT) [35, theorem 11.2.5] states that as K tends to infinity it holds true that[4]:

$$\frac{\displaystyle\sum_{k=1}^{K}\sum_{l=2}^{64}\Lambda^{\rm lr}(u_{k,l}) - E_{\mathcal{H}_i}(\theta_{k,l})}{\left(\displaystyle\sum_{k=1}^{K}\sum_{l=2}^{64}V_{\mathcal{H}_i}(\theta_{k,l})\right)^{1/2}} \xrightarrow{d} \mathcal{N}(0,1)\ , \ i = \{0,1\}\,, \tag{8}$$

where $\xrightarrow{d}$ represents the convergence in distribution and $\mathcal{N}(0,1)$ is the standard normal distribution, *i.e.* with zero mean and unit variance.

[4]Note that we refer to the Lindeberg's CLT, whose conditions are easily verified in our case, because the random variable are independent but are not i. i. d.

This theorem is of crucial interest to establish the statistical properties of the proposed test [38, 7, 37, 22]. In fact, once the moments have been calculated under both $\mathcal{H}_i, i = \{0,1\}$, one can normalise under hypothesis $\mathcal{H}_0$ the LR $\Lambda^{\mathrm{lr}}(\mathbf{U})$ as follows:

$$\overline{\Lambda}^{\mathrm{lr}}(\mathbf{U}) = \frac{\Lambda^{\mathrm{lr}}(\mathbf{U}) - \sum_{k=1}^{K}\sum_{l=2}^{64} E_{\mathcal{H}_0}(\theta_{k,l})}{\left(\sum_{k=1}^{K}\sum_{l=2}^{64} V_{\mathcal{H}_0}(\theta_{k,l})\right)^{1/2}}, \\ = \frac{\sum_{k=1}^{K}\sum_{l=2}^{64} \Lambda^{\mathrm{lr}}(u_{k,l}) - E_{\mathcal{H}_0}(\theta_{k,l})}{\left(\sum_{k=1}^{K}\sum_{l=2}^{64} V_{\mathcal{H}_0}(\theta_{k,l})\right)^{1/2}}. \quad (9)$$

Since this essentially consists in adding a deterministic value and scaling the LR, this operation of normalisation preserves the optimality of the LRT. It is thus straightforward to define the normalised LRT with $\overline{\Lambda}^{\mathrm{lr}}(\mathbf{U})$ by:

$$\overline{\delta}^{\mathrm{lr}}(\mathbf{U}) = \begin{cases} \mathcal{H}_0 & \text{if } \overline{\Lambda}^{\mathrm{lr}}(\mathbf{U}) < \overline{\tau}^{\mathrm{lr}} \\ \mathcal{H}_1 & \text{if } \overline{\Lambda}^{\mathrm{lr}}(\mathbf{U}) \geq \overline{\tau}^{\mathrm{lr}}. \end{cases} \quad (10)$$

It immediately follows from Lindeberg's CLT (8) that $\overline{\Lambda}^{\mathrm{lr}}(\mathbf{U})$ asymptotically follows, as K tends to infinity, the normal distribution $\mathcal{N}(0,1)$. Hence, it is immediate to set the decision threshold that guarantee the prescribed false alarm probability:

$$\overline{\tau}^{\mathrm{lr}} = \Phi^{-1}\left(1 - \alpha_0\right), \quad (11)$$

where Φ and Φ^{-1} respectively represent the cumulative distribution function (cdf) of the standard normal distribution and its inverse. Similarly, denoting

$$m_i = \sum_{k=1}^{K}\sum_{l=2}^{64} E_{\mathcal{H}_i}(\theta_{k,l}); \sigma_i^2 = \sum_{k=1}^{K}\sum_{l=2}^{64} V_{\mathcal{H}_i}(\theta_{k,l}), i = \{0,1\},$$

it is also straightforward to establish the detection function of the LRT given by:

$$\beta_{\overline{\delta}^{\mathrm{lr}}} = 1 - \Phi\left(\frac{\sigma_0}{\sigma_1}\Phi^{-1}\left(1-\alpha_0\right) + \frac{m_0 - m_1}{\sigma_1}\right). \quad (12)$$

Equations (11) and (12) emphasize the main advantage of normalising the LR as described in relation (9): it allows to set any of threshold that guarantee a false alarm probability independently from any distribution parameters and, this is particularly crucial because digital images are heterogeneous, their properties vary for each image. Second, the normalisation allows to easily establish the detection power which again, is achieved, for any distribution parameters and hence, for any inspected image.

3.3 Application with Laplacian distribution

In the case of Laplacian distribution, the framework of hypothesis testing theory has been applied for the steganalysis of JSteg in [22] in which the moments of LR are calculated under the two following assumptions: 1) all the DCT coefficients from the same subband are i. i. d. and 2) the expectation of each DCT coefficient is zero.
The continuous Laplacian distribution has the following probability density function:

$$f_{\mu,b}(x) = \frac{1}{2b}\exp\left(-\frac{|x-\mu|}{b}\right) \quad (13)$$

where $\mu \in \mathbb{R}$, sometimes referred to as the location parameter, corresponds to the expectation, and $b > 0$ is the so-called scale parameter. During the compression of JPEG images, the DCT coefficients are quantised. Hence, let us defined the discrete Laplacian distribution by the following pmf, see details in Appendix A:

$$f_{\mu,b}[k] \stackrel{def.}{=} \mathbb{P}\Big[x \in [\Delta(k - 1/2), \Delta(k + 1/2)[\Big] \\ = \begin{cases} \exp\left(-\frac{|\Delta k - \mu|}{b}\right)\sinh\left(\frac{\Delta}{2b}\right) & \text{if } \frac{\mu}{\Delta} \notin [k-1/2; k+1/2[\\ 1 - \exp\left(-\frac{\Delta}{2b}\right)\cosh\left(-\frac{\Delta(k)-\mu}{2b}\right) & \text{otherwise} \end{cases} \quad (14)$$

where Δ is the quantization step.
From the expression of the discrete Laplacian distribution (14) and from the expression of LR (7), one can express the LR for the detection of JSteg under the assumption that DCT coefficients follow a Laplacian distribution, as follows (see Appendix B):

$$\Lambda^{\mathrm{lr}}_{\mu,b}[k] = \log\left(1 - \frac{R}{2} + \frac{R}{2}\exp\left[\frac{\Delta}{b}\operatorname{sign}(\Delta k - \mu)(k - \bar{k})\right]\right). \quad (15)$$

where DCT coefficient k is referred as $u_{k,l}$ in Eq. (7). It can be noted that this expression (15) of the LR is almost the same as the one obtained in [22] assuming that all DCT coefficients have a zero-mean, only the sign term $\operatorname{sign}(\Delta k - \mu)$ becomes $\operatorname{sign}(k)$ when assuming a zero-mean. It should also be noted that the log-LR equals 0 for every DCT coefficient whose value is 0 or 1 because the JSteg algorithm does not embed hidden data in those coefficients. In the present paper, the moments of the LR (15) are not analytically established, the reader interested can refer to [22].

4 Proposed approach for estimating the nuisance parameters in practice

4.1 Estimation of expectation of each DCT coefficient

As already explained, most of statistical models of DCT coefficients assume that within a subband the coefficients are i. i. d.. However, as illustrated in Figure 1b and 2 this assumption is doubtful in practice. Another way to explain why the DCT coefficients may not be i. i. d. is to consider a block of 8×8 pixels in spatial domain, say the first, $\mathbf{z} = z_{i,j}\,, i \in \{1,\dots,8\}\,, j \in \{1,\dots,8\}$. The value of those pixels can be decomposed as:

$$z_{i,j} = x_{i,j} + n_{i,j},$$

where $x_{i,j}$ is a deterministic value that represents the expectation of pixel at location (i,j) and $n_{i,j}$ is the realisation of a random variable representing all noises corrupting the inspected image. Clearly, this decomposition can be done for the whole block $\mathbf{z} = \mathbf{x}+\mathbf{n}$, where $\mathbf{x} = \{x_{i,j}\}$ and $\mathbf{n} = \{n_{i,j}\}$. Since the DCT transformation is linear the DCT coefficient of any block may be expressed as :

$$\begin{aligned} &\mathrm{DCT}(\mathbf{z}) = \mathbf{D}^T\mathbf{z}\mathbf{D} = \mathbf{D}^T(\mathbf{x}+\mathbf{n})\mathbf{D} \\ =&\mathbf{D}^T\mathbf{x}\mathbf{D} + \mathbf{D}^T\mathbf{n}\mathbf{D} = \mathrm{DCT}(\mathbf{x}) + \mathrm{DCT}(\mathbf{n}), \end{aligned} \qquad (16)$$

where DCT represents the DCT transform and $\mathbf{D}$ is the change of basis matrix from spatial to DCT basis, often referred as the DCT matrix.

It makes sense to assume that the expectation of the noise component $\mathbf{n}$ has a zero-mean in the spatial and in the DCT domain. On the opposite, it is difficult to justify that the DCT of pixels' expectation $\mathbf{x}$ should necessary be around zero. Actually, this assumption holds true if and only if the expectation is the same for of all the pixels from a block: $\forall i \in \{1,\dots,8\}\,, \forall j \in \{1,\dots,8\}\,, x_{i,j} = x$, see [36, 37, 39] for details.

On the opposite, in the paper, it is mainly aimed at estimating the expectation of each DCT coefficient. To this end, it is proposed to decompress a JPEG image $\mathbf{V}$ into the spatial domain to obtain $\mathbf{Z}$, then to estimate the expectation of each pixel in the spatial domain $\widehat{\mathbf{Z}}$ by using a denoising filter. Then this denoised image $\widehat{\mathbf{Z}}$ is transformed back into the DCT domain to finally obtain the estimated value of all DCT coefficients, denoted $\widehat{\mathbf{V}} = \{\hat{v}_{i,j}\}\,, i \in \{1,\dots,I\}\,, j \in \{1,\dots,J\}$. Several methods have been tested to estimate the expectation of pixels in the spatial domain $\widehat{\mathbf{Z}}$, namely, the BM3D collaborative filtering [40], K-SVD sparse dictionary learning [41], non-local weighted averaging method from NL-means [42] and the wavelet denoising filter [43]. The codes used for the methods [40, 41, 42] have been downloaded from the Image Processing On-Line website[5]. The codes used for the method [43] have been downloaded from DDE [6].

4.2 A local estimation of b

In addition, the proposed model also assumes that the scale parameter $b_{k,l}$ is different for each DCT coefficient. The estimation of this parameter, for each DCT coefficient, is based on the WS Jpeg method to locally estimate the variance; that is, for coefficients $v_{i,j}$, it simply consists of the sample variance of the DCT coefficients of the same subband from neighbouring blocks:

$$\widehat{\sigma}^2_{i,j} = \frac{1}{7}\sum_{\substack{s=-1\\(s,t)\neq(0,0)}}^{1}\sum_{t=-1}^{1}\left(v_{i+8s,j+8t} - \bar{v}_{i,j}\right)^2, \qquad (17)$$

where $\bar{v}_{i,j}$ is the sample mean: $\frac{1}{8}\sum_{\substack{s=-1\\(s,t)\neq(0,0)}}^{1}\sum_{t=-1}^{1} v_{i+8s,j+8t}$.

Let us recall that the Maximum Likelihood Estimation (MLE) of the scale parameter of Laplacian distribution from realisations $x1,\dots,x_N$ is given by $\hat{b} = N^{-1}\sum_{n=1}^{N}|x_n-\mu|$. The local estimation of the scale parameter it is proposed to use in this paper is given by:

$$\hat{b}_{i,j} = \frac{1}{8}\sum_{\substack{s=-1\\(s,t)\neq(0,0)}}^{1}\sum_{t=-1}^{1}\left|v_{i+8s,j+8t} - \hat{v}_{i+8s,j+8t}\right|, \qquad (18)$$

where $\hat{v}_{i+8s,j+8t}$ is the estimation of expectation of each DCT coefficient by using denoising filter previously defined. As in the WS Jpeg algorithm, this approach raises the problem of scale parameter estimation for blocks located on the sides of the image. In the present paper, as in the WS Jpeg method, it is proposed not to use those blocks in the test.

4.3 A channel selection to improve the method

Inspired by the channel selection algorithms (See [28, 29]), it is proposed to improve our detector with a weighting factor (WF). In practice, WF is generated from the quantized and rounded "residual noise", which is calculated by the following steps:

1 By uncompressing the JPEG format image, we obtain the intensity value of a JPEG image in the spatial domain.

2 By using a denoising filter, we extract the raw "residual noise" in the spatial domain.

[5]Image Processing On-Line journal is available at: `http://www.ipol.im`

[6]Source codes are available at: `http://dde.binghamton.edu`

Table 1: **Ratio (%) comparison before and after embedding.**

	Inspected images index										
	No.1	**No.2**	**No.3**	**No.4**	**No.5**	**No.6**	**No.7**	**No.8**	**No.9**	**No.10**	**On average**
Cover_N	0.23	0.17	0.56	0.61	0.21	0.03	0.87	0.41	1.23	0.33	0.63
Stego_N	0.23	0.17	0.56	0.62	0.21	0.04	0.88	0.42	1.22	0.34	0.64
Cover_S	0.98	1.01	1.06	1.03	0.90	1.07	1.02	0.93	1.26	1.03	7.45
Stego_S	0.98	1.00	1.06	1.03	0.89	1.07	1.03	0.90	1.27	1.03	7.52
Cover_D	1.12	0.81	2.46	2.49	0.80	0.08	5.07	2.42	7.56	0.34	1.44
Stego_D	4.48	2.85	7.63	7.60	2.27	1.07	17.1	7.92	20.5	1.04	4.95
Similarity	89.5	91.5	94.2	93.0	80.7	80.7	93.9	93.7	93.3	93.9	92.8

3 By using DCT transformation, we transform the raw "residual noise" from the spatial to the frequency domain.
4 By using quantization table, we can obtain the quantized "residual noise".
5 By rounding the quantized "residual noise" in the frequency domain, the quantized and rounded "residual noise" is obtained.
6 If a quantized and rounded "residual noise" takes zero, WF equals 0; If not, WF equals 1.

Thus, based on our proposed WF, it is proposed to categorize "residual noise" set into two sub-sets: "non-zero" sub-set and "zero" sub-set. To verify the effectiveness of our improved algorithm, it is proposed to randomly choose ten exemplary images which are compressed to JPEG format images with QF = 70, embedding rate $R = 0.05$. Also, all the images of BossBase database [27] are used for computing the average value. Table 1 gives the statistical ratio of the data in which the annotations of the table are as followed:

- **Cover_N**: denotes the ratio of "non-zero" sub-set to "residual noise" set of a cover image.
- **Stego_N**: denotes the ratio of "non-zero" sub-set to "residual noise" set of a stego image.
- **Cover_S**: denotes the standard deviation of "residual noise" set from a cover image.
- **Stego_S**: denotes the standard deviation of "residual noise" set from a stego image.
- **Cover_D**: denotes the ratio of the DCT coefficients used by JSteg in "non-zero" sub-set to the DCT coefficients used by JSteg in "residual noise" set from a cover image .
- **Stego_D**: denotes the ratio of the DCT coefficients used by JSteg in "non-zero" sub-set to the DCT coefficients used by JSteg in "residual noise" set from a stego image .
- **Similarity**: denotes the ratio of the same position in "non-zero" sub-set before and after embedding.

In our proposed statistical test, the number of the selected coefficients for the detection should be kept very close before and after embedding. As Table 1 illustrated, the ratio of **Cover_N** and **Stego_N** basically remains the same before and after embedding, which reveals the proportion of the coefficients used for the test nearly the same. Similarly, the ratio of **Cover_S** and **Stego_S** can also verify our assumption that the statistical number nearly remains the same before and after embedding. The ratio of **Similarity** which is kept at the high value signifies most of "residual noise" are chosen at the same position. Then the only difference is the comparison between **Cover_D** and **Stego_D**. It is should be noted that if all DCT coefficients used by JSteg are included in "non-zero" sub-set, then the ratio equals 100%. It is observed that only a few of DCT coefficients used by JSteg algorithm is included in "non-zero" sub-set. Nevertheless, after embedding, the ratio of **Stego_D** is largely improved, compared with the ratio of **Cover_D**. It can be assumed that by using a WF, more "residual noise" from the embedding positions are counted. Besides, prior to embedding secret information, we never know which position will be embedded, the very low ratio of **Cover_D** is reasonable.

By investigating the "non-zero" and "zero" sub-set, although we can not capture all the embedding positions in the DCT domain, it is totally enough to detect the JSteg steganography. Besides, all the coefficients in "zero" sub-set are not counted in our proposed test. On average, for a cover image with the size of 512×512, 0.63% of the coefficients are kept to compute the test; 0.64% of the coefficients from a stego image are used. As the embedding rate $R = 0.05$, it is obvious that most of DCT coefficients remain the same before and after embedding. Thus, it is not necessary to compute these values. Furthermore, the LR values of these DCT coefficients without embedding any information probably mask or disturb LR from DCT coefficients with JSteg embedding.

4.4 Design of proposed test

In Section 3 the framework of hypothesis testing theory has been presented assuming that distribution parameters are known for each DCT coefficient. To design a practical test, a usual solution consists in replacing the unknown parameter by its ML estimation. This leads to the construction of a Generalised LRT. A similar construction is adopted in this paper, using the *ad hoc* estimators presented at the beginning of section 4, instead of using the ML method to estimate the distribution parameters of each DCT coefficient. The proposed test is thus defined as:

$$\widehat{\delta}(\mathbf{U}) = \begin{cases} \mathcal{H}_0 \text{ if } \widehat{\Lambda}(\mathbf{U}) = \sum_{k=1}^{K}\sum_{l=2}^{64} \widehat{\Lambda}_{cs}(u_{k,l}) < \widehat{\tau}, \\ \mathcal{H}_1 \text{ if } \widehat{\Lambda}(\mathbf{U}) = \sum_{k=1}^{K}\sum_{l=2}^{64} \widehat{\Lambda}_{cs}(u_{k,l}) \geq \widehat{\tau}, \end{cases} \tag{19}$$

where the channel selection decision statistic $\widehat{\Lambda}_{cs}(u_{k,l}) = \widehat{\Lambda}(u_{k,l}) \cdot w_{k,l}$ for a single DCT coefficient is given and a weighting factor $w_{k,l}$ selects the DCT channel. Next, let us study the $\widehat{\Lambda}(u_{k,l})$ to verify the effectiveness of our proposed test.

To verify our improvement based on the Laplacian test, see [22], it is proposed to consider the weighing factor $w_{k,l}$ as a constant equal to 1. The scale parameter $\widehat{b}$ is estimated by using MLE and the location parameter is ignored (See details in [22]). The LR is given by:

$$\widehat{\Lambda}(u_{k,l}) = \log\left(1 + \frac{R}{2} + \frac{R}{2}\exp\left[\frac{\Delta}{\widehat{b}}\operatorname{sign}(\Delta k)(k-\bar{k})\right]\right). \tag{20}$$

The first improvement of the previous LR is the consideration of the location parameter $\widehat{\mu}_{k,l}$ (See 4.1). The new LR is designed by:

$$\widehat{\Lambda_1}(u_{k,l}) = \log\left(1 + \frac{R}{2} + \frac{R}{2}\exp\left[\frac{\Delta}{\widehat{b}}\operatorname{sign}(\Delta k - \widehat{\mu}_{k,l})(k-\bar{k})\right]\right). \tag{21}$$

The second improvement is the estimation of the scale parameter $\widehat{b}_{k,l}$ (See 4.2) and ignore the location parameter. The LR is designed by:

$$\widehat{\Lambda_2}(u_{k,l}) = \log\left(1 + \frac{R}{2} + \frac{R}{2}\exp\left[\frac{\Delta}{\widehat{b}_{k,l}}\operatorname{sign}(\Delta k)(k-\bar{k})\right]\right). \tag{22}$$

The third improvement is to give the assumption that DCT coefficients are i. i. d.. The scale parameter $\widehat{b}_{k,l}$ and the location parameter $\widehat{\mu}_{k,l}$ of the distribution are estimated separately by using our proposed algorithms of 4.1 and 4.2.

$$\widehat{\Lambda_3}(u_{k,l}) = \log\left(1 + \frac{R}{2} + \frac{R}{2}\exp\left[\frac{\Delta}{\widehat{b}_{k,l}}\operatorname{sign}(\Delta k - \widehat{\mu}_{k,l})(k-\bar{k})\right]\right). \tag{23}$$

Moreover, it is proposed to explore the effectiveness of introducing a weighing factor $w_{k,l}$ which is defined as:

$$w_{k,l} = \begin{cases} 0 & \text{if} \quad \Delta k - \widehat{\mu}_{k,l} \in (-0.5, 0.5) \\ 1 & \text{otherwise.} \end{cases} \tag{24}$$

The last LR is obtained by multiplying (23) by $w_{k,l}$:

$$\widehat{\Lambda}_{cs}(u_{k,l}) = \widehat{\Lambda_3}(u_{k,l}) w_{k,l} \tag{25}$$

It is should be noted that (20) is the algorithm from [22]. In Section 5, the specific comparison of the detectors is presented. In order to have a normalised decision statistic for the whole image, $\widehat{\Lambda}(\mathbf{U})$ is defined as:

$$\begin{aligned} \widehat{\Lambda}(\mathbf{U}) &= \frac{1}{S_L}\sum_{k=1}^{K}\sum_{l=2}^{64}\widehat{\Lambda}_{cs}(u_{k,l}) - E_{\mathcal{H}_0}(\widehat{\mu}_{k,l}, \widehat{b}_{k,l}) \\ \text{with} \quad S_L^2 &= \sum_{k=1}^{K}\sum_{l=2}^{64} V_{\mathcal{H}_0}(\widehat{\mu}_{k,l}, \widehat{b}_{k,l}). \end{aligned} \tag{26}$$

4.5 Comparison with prior-art

The WS Jpeg, as well as the WS for spatial domain, is based on the underlying assumption that the observations follow a Gaussian distribution. As recently shown [6, 8], the WS implicitly assumes that the quantization step is negligible. Let us rewrite the LR test for JSteg detection based on a Gaussian distribution model of DCT coefficients. Let X be a random variable following a quantized Gaussian distribution. Exploiting the assumption that the quantization step is negligible compared to noise standard deviation allows the writing of:

$$\begin{aligned} \mathbb{P}[X = k] &= \int_{\Delta(k-1/2)}^{\Delta(k+1/2)} \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)dx \\ &\approx \frac{\Delta}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(\Delta k-\mu)^2}{2\sigma^2}\right) \end{aligned} \tag{27}$$

Putting this expression of the pmf under hypothesis $\mathcal{H}_0$ into the LR (2), and assuming that the quantization step is negligible compared to the noise standard deviation, $\Delta << \sigma$, it is immediate to obtain the following expression of the LR under the assumption of Gaussian distribution of DCT coefficient

$$\log\left(1+\frac{R}{2}+\frac{R}{2}\frac{\exp\left(-\frac{(\Delta\bar{k}-\mu)^2}{2\sigma^2}\right)}{\exp\left(-\frac{(\Delta k-\mu)^2}{2\sigma^2}\right)}\right)$$

$$\approx \frac{R\Delta}{2\sigma^2} \quad (k-\bar{k}) \quad (\Delta k-\mu) \qquad (28)$$

$$= \underbrace{}_{w_\sigma} \quad \underbrace{}_{\pm 1} \quad \underbrace{}_{(\Delta k-\mu)}$$

see details in Appendix C.
This expression highlights the well known fact the WS consists in fact of three terms: 1) the term w_σ which is a weight so that pixels or DCT coefficients with highest variance have a smallest importance, 2) the term $(k-\bar{k}) = \pm 1$ according the LSB of k and 3) the term $(\Delta k-\mu)$.

In comparison, the expression of the LR for a Laplacian distribution model (15), as well as the expression of the proposed test with estimates (21) can be approximated by (See details in Appendix B):

$$\frac{R\Delta}{2b} \quad (k-\bar{k}) \quad \mathrm{sign}(\Delta k-\mu) \qquad (29)$$

$$= \underbrace{}_{w_b} \quad \underbrace{}_{\pm 1} \quad \underbrace{}_{\mathrm{sign}(\Delta k-\mu)}$$

which is also made of three terms; the two first are roughly similar to the two first terms of the WS : 1) the term w_b is a weight so that DCT coefficients with highest "scale" b have a smallest importance, note that the variance is proportional to b^2, 2) the term $(k-\bar{k}) = \pm 1$ according to the LSB of k. However, in the expression of the LR based on the Laplacian model the term $(\Delta k-\mu)$ of the WS is replaced with its sign. This shows that the statistical tests based on Laplacian model and based on Gaussian model are essentially similar.

5 Numerical simulations

5.1 Results on simulated images

One of the main contributions of this paper is to show that the hypothesis testing theory can be applied in practice to design a statistical test with known statistical properties for JSteg steganalysis.

To verify the sharpness of the theoretically established results, we generate 1000 sets of 4000 random variable (a Monte-Carlo simulation) following the Laplacian distribution, where $R = 0.05$, $\mu = 0$ and b distributed from 1 to 10 with a step of 0.5. Then, the expectation and variance values are calculated empirically and theoretically. As shown in Figure 5, the empirically calculated moments are almost equal to the analytically established ones.

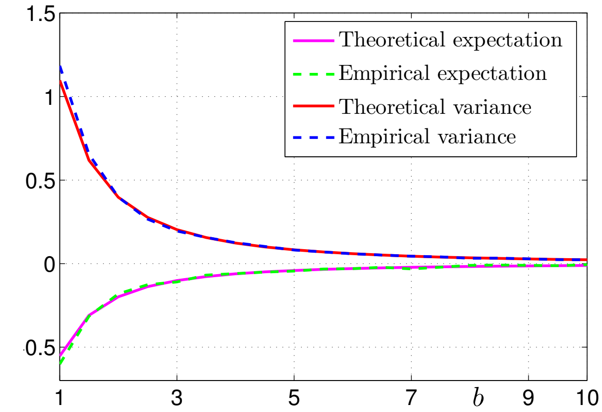


Figure 5: **Expectation m_0 and variance σ_0^2 as a function of the scale parameter b theoretically and empirically.**

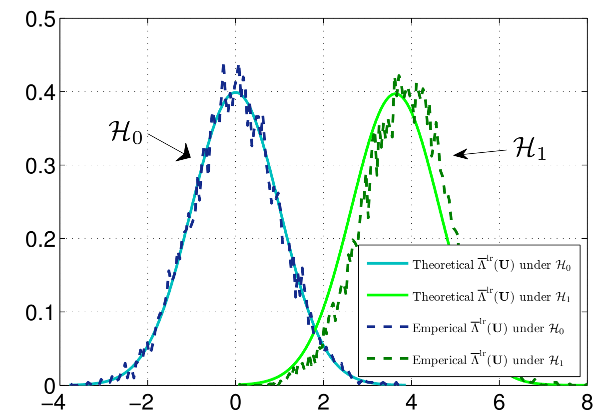


Figure 6: **Comparison between empirical and theoretical distribution of $\overline{\Lambda}^{\mathrm{lr}}(\mathbf{U})$.**

Subsequently, to verify the effectiveness of the established LRT $\overline{\delta}^{\mathrm{lr}}(\mathbf{U})$, again, a Monte-Carlo simulation is performed by repeating 10000 times using a vector 64×4096 following the Laplacian distribution, in which the scale parameter is selected arbitrarily as 3 and the location parameter 0. Under the hypothesis $\mathcal{H}_0$ and $\mathcal{H}_1$ respectively, Figure 6 presents the comparison between empirical and theoretical distribution of $\overline{\Lambda}^{\mathrm{lr}}(\mathbf{U})$. The results highlight the validity of the proposed test (10).

Figure 7 gives the comparison between the empirical and theoretical FAR α_0 respectively of the test (10). This particularly demonstrates that two curves are very close. Figure 8 offers the Receiver Operating Characteristic (ROC) comparison, that is the detection power $\beta_{\overline{\delta}^{\mathrm{lr}}}$ as a function of FAR α_0, of both empirical and theoretical established results in 11 and 12.

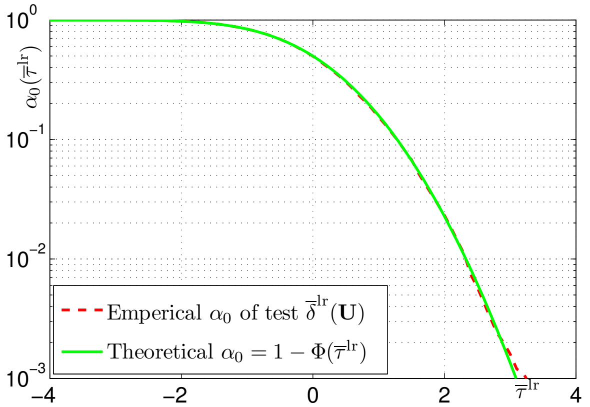


Figure 7: **FAR α_0 as a function of the threshold $\overline{\tau}^{\mathrm{lr}}$.**

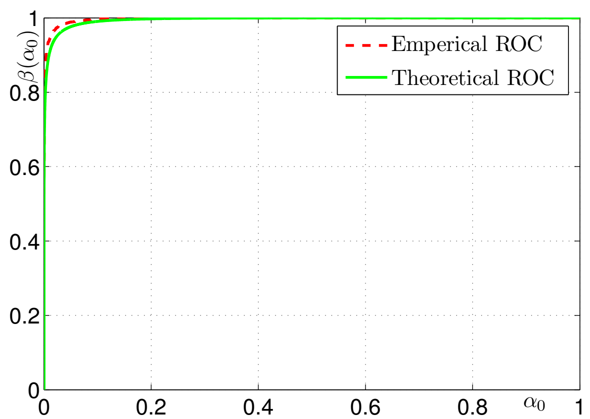
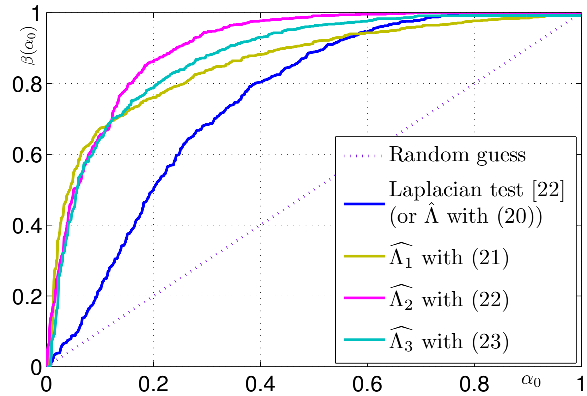
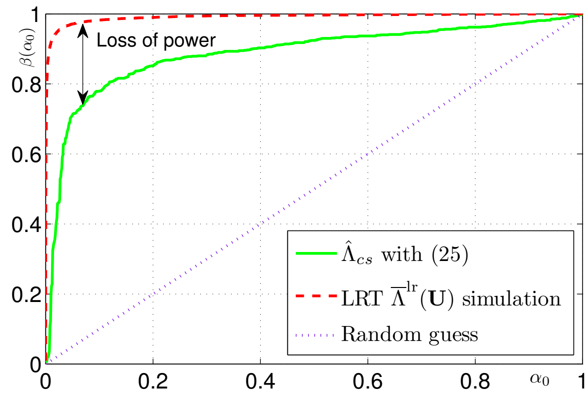


Figure 8: **Detection power $\beta_{\overline{\delta}^{\mathrm{lr}}}$ as a function of FAR α_0 (ROC curve).**



(a) Comparison of detection performance without channel selection (constant weighting factor $w_{k,l} = 1$).



(b) Comparison of detection performance with channel selection (weighting factor $w_{k,l}$ from (24)).

Figure 9: **ROC curves comparison, detection power as a function of FAR α_0.**

5.2 Results on real images

Another contribution of this paper is to design the optimal test with estimated parameters to break JSteg algorithm in practical case.

First, let us investigate our proposed detectors (21)-(23). It is proposed to perform a numerical simulation over the 1000 images from BossBase [27] which have been compressed in JPEG with quality factor 70. The payload, or embedding rate, R is set at 0.05 for JSteg algorithm. For fairly comparison with the detector from [22], it first shows the improvement provided by the proposed model with $w_{k,l} = 1$. As Figure 9a illustrates, all the proposed detectors outperforms $\widehat{\Lambda}(u_{k,l})$ (20) proposed by [22]. Morvover, in the following investigation, it is proposed to use $\widehat{\Lambda}_{cs}(u_{k,l})$ (25). Then, it is proposed to give the performance of this detector on 1000 simulated images in which a DCT subband is generated by strictly following the Laplacian distribution (See Figure 9b). Then a comparison with simulations of the LR test shows the loss of power due to the estimation of expectation and scale parameters. It is should be noted that in all our proposed detectors in this paper, $\widehat{\Lambda}(u_{k,l})$ (23) with $w_{k,l}$ (24) performs best. Thus, it is proposed to use it as our optimal steganalyser for competing with the state-of-the-art JSteg detectors. It is should be emphasized that in Figure 9, the wavelet denoising filter [43] is used for estimating the location parameter $\widehat{\mu}_{k,l}$ (See 4.1).

To verify the relevance of the proposed methodology, it is proposed to compare the proposed statistical test with two other detectors. The first chosen competitor is the statistical test proposed in [22] as it is also based on Laplacian model but does not take into account the distribution parameters as nuisance parameters; it considers that DCT coefficients are i. i. d. , following a Laplacian distribution with zero-mean. The comparison with this test is meaningful as it allows us to measure how much the detection performance is improved by removing the assumption that the DCT coefficients of each subband are i. i. d. . The second chosen competitor is the WS [15] due to its similarity with the proposed statistical test, see details in Section 4.5.

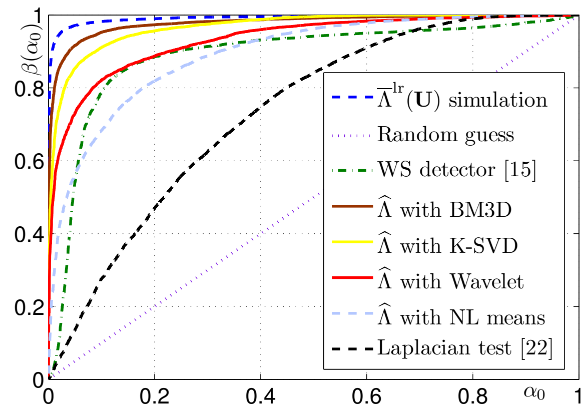
For a large scale verification, it is proposed to use the BOSS database, made of 10 000 grayscale images of size 512×512 pixels, used with payload $R = 0.05$. Prior to our experiments, the images have been compressed in JPEG using the linux command convert which uses the standard quantization table. Note also that all the JSteg steganography was performed using a Matlab source code we developed based on Phil Sallee's Jpeg Toolbox[7]. Three denoising methods have been tested to estimate the expectation of each DCT coefficient, namely the K-SVD, the BM3D, the wavelet denoising algorithms.

Figure 10 shows the detection performances obtained over the BOSS database compressed with quality factor (QF) 70. The detection performances are shown as ROC curves, that is the detection power is plotted as a function of false-alarm probability. The Figure 10a particularly emphasizes that the statistical test based on the Laplacian model does not perform well while the proposed methodology which takes into account the Laplacian distribution parameters as nuisance parameters allows us to largely improve the performance. Similarly the WS detector achieves overall good detection performance. However, it can be shown on Figure 10b, which presents the same results using a logarithmic scale, that for low false-alarm probabilities, the performance of the WS significantly decreases. On the opposite, the proposed statistical test still performs well.

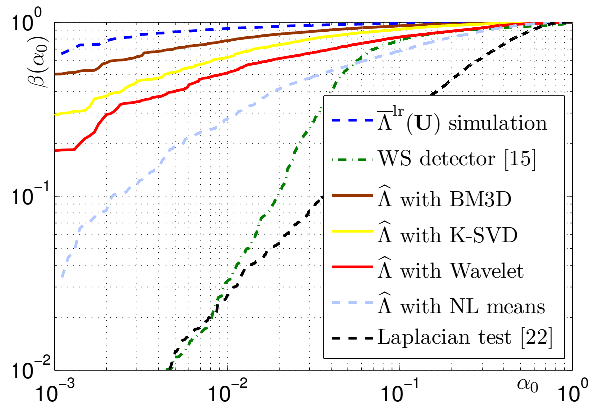
Among the four denoising algorithms that have been tested, the BM3D achieves the best performance but it can be observed on Figure 10 that the performance obtained using the K-SVD and using the wavelet denoising methods are also very good. The performance of NL means method is comparable with WS detector [15].

To extend the results previously presented, a similar test has been performed over the BOSS database using the quality factor 85. The detection performance obtained by the proposed test and by the competitors are presented in Figure 11. Again, this figure shows that based on the Laplacian model, the statistical test assuming that DCT coefficients of a subband are i. i. d. has an unsatisfactory performance. It can also be noted that even though the WS performs slightly better for low false-alarm probability, compared to the results obtained with quality factor 70, it performs much worse than the proposed statistical test.

[7] Phil Sallee's Jpeg Toolbox is available at : `http://dde.binghamton.edu/download/jpeg_toolbox.zip`



(a) Comparison of detection performance for BOSS database with quality factor 70 (linear scale).



(b) Comparison of detection performance for BOSS database with quality factor 70 (logarithmic scale).

Figure 10: **Comparison of detection performance for BOSS database with quality factor 70.**

6 Conclusion and future works

This paper aims at improving the optimal detection of data hidden within the DCT coefficients of JPEG images. Its main originality is that the usual Laplacian model is used as a statistical model of DCT coefficients but, opposed to what is usually proposed, it is not assumed that all DCT coefficients from a subband arei. i. d. . This leads us to consider the Laplacian distribution parameters, namely the expectation e and the scale parameter b, as nuisance parameters as they have no interest for the detection of hidden data, but must be carefully taken into account to design an efficient statistical test. Numerical results show that by estimating those nuisance parameters, the Laplacian model allows the designing of an accurate statistical test which outperforms the WS. The comparison with the optimal detector based on the Laplacian model and on the assumption that all DCT coefficients of a

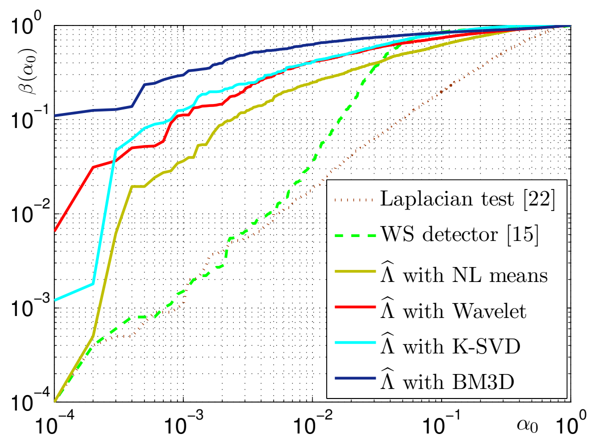


Figure 11: **Comparison of detection performance for BOSS database with quality factor 85 (logarithmic scale).**

subband are i. i. d. shows the relevance of the proposed approach.

A possible future work would be to apply this approach with state-of-the-art statistical model of DCT coefficients, such as the Generalized Gaussian or the Generalized Gamma model. This could provide improvements in the detection performance at the cost of a higher complexity.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
TQ carried out the main research of this work and drafted the manuscript. FR, RC and CZ helped to modify the manuscript. All authors read and approved the final manuscript.

Acknowledgements
The Matlab codes will be published upon paper acceptance. The work of F. Retraint, R. Cogranne and C. Zitzmann is funded by Troyes University of Technology (UTT) strategic program COLUMBO. The PhD thesis of Tong Qiao is funded by the China Scholarship Council (CSC) program.

Appendix A: Quantized Laplacian PMF
Let X be a Laplacian random variable with expectation μ and variance b. Its pdf is thus, see (13):

$$f_{\mu,b}(x) = \frac{1}{2b} \exp\left(-\frac{|x-\mu|}{b}\right),$$

and a straightforward calculation shows that its cdf is given by:

$$F_{\mu,b}(x) = \frac{1}{2} + \frac{1}{2}\,\mathrm{sign}(x-\mu)\left(1 - \exp\left(-\frac{|x-\mu|}{b}\right)\right), \tag{30}$$

$$= \begin{cases} \frac{1}{2}\exp\left(\frac{x-\mu}{b}\right) & \text{if } x < \mu, \\ 1 - \frac{1}{2}\exp\left(-\frac{x-\mu}{b}\right) & \text{if } x \geq \mu. \end{cases} \tag{31}$$

Now consider the result from quantization of this random variable $Y = \lfloor X/\Delta \rfloor$, it is immediate to establish the pmf of this random variable. Let us first consider the case $\Delta(k+1/2) < \mu$ (due to the symmetry of Laplacian pdf, the case $\Delta(k-1/2) > \mu$ is treated similarly).

The pmf of Y is given by:

$$\begin{aligned} \mathbb{P}[Y=k] &= \mathbb{P}[\Delta(k-{}^1/_2) \leq X < \Delta(k+{}^1/_2)], \\ &= \frac{1}{2}\exp\left(\frac{\Delta(k+{}^1/_2)-\mu}{b}\right) \\ &\quad - \frac{1}{2}\exp\left(\frac{\Delta(k-{}^1/_2)-\mu}{b}\right), \\ &= \frac{1}{2}\exp\left(\frac{\Delta k-\mu}{b}\right)\exp\left(\frac{\Delta}{2b}\right) \\ &\quad - \frac{1}{2}\exp\left(\frac{\Delta k-\mu}{b}\right)\exp\left(\frac{-\Delta}{2b}\right), \\ &= \exp\left(\frac{\Delta k-\mu}{b}\right)\sinh\left(\frac{\Delta}{2b}\right), \end{aligned}$$

Applying similar calculations for case $\Delta(k-1/2) > \mu$, one gets:

$$\mathbb{P}[Y=k] = \exp\left(-\frac{|\Delta k-\mu|}{b}\right)\sinh\left(\frac{\Delta}{2b}\right), \tag{32}$$

which corresponds to the pmf given in Eq. (14). The case $\Delta(k-1/2) < \mu < \Delta(k+1/2)$ is treated similarly.

Appendix B: Log-Likelihood Ratio Calculation
By putting the expression of quantised Laplacian pmf (32) into the expression of the LR (7), it is immediate to write:

$$\Lambda^{\mathrm{lr}}(u_{k,l}) = \log\left(1 - \frac{R}{2} + \frac{R}{2}\,\frac{\exp\left(-\frac{|\Delta\bar{k}-\mu|}{b}\right)\sinh\left(\frac{\Delta}{2b}\right)}{\exp\left(-\frac{|\Delta k-\mu|}{b}\right)\sinh\left(\frac{\Delta}{2b}\right)}\right).$$

Let us study the term:

$$
\begin{aligned}
&\frac{\exp\left(-\frac{|\Delta\bar{k}-\mu|}{b}\right)\sinh\left(\frac{\Delta}{2b}\right)}{\exp\left(-\frac{|\Delta k-\mu|}{b}\right)\sinh\left(\frac{\Delta}{2b}\right)} = \frac{\exp\left(-\frac{|\Delta\bar{k}-\mu|}{b}\right)}{\exp\left(-\frac{|\Delta k-\mu|}{b}\right)},\\
=&\frac{\exp\left(-\frac{|\Delta k+\Delta(\bar{k}-k)-\mu|}{b}\right)}{\exp\left(-\frac{|\Delta k-\mu|}{b}\right)},\\
=&\frac{\exp\left(-\frac{|\Delta k-\mu|}{b}\right)\exp\left(\frac{\Delta\,\mathrm{sign}(\Delta k-\mu)(k-\bar{k})}{b}\right)}{\exp\left(-\frac{|\Delta k-\mu|}{b}\right)},\\
=&\exp\left(\frac{\Delta\,\mathrm{sign}(\Delta k-\mu)(k-\bar{k})}{b}\right).
\end{aligned}
\tag{33}
$$

From this Eq. (33), it is immediate to establish the expression (15):

$$
\log\left(1-\frac{R}{2}+\frac{R}{2}\exp\left(\frac{\Delta\,\mathrm{sign}(\Delta k-\mu)(k-\bar{k})}{b}\right)\right).
$$

By using a Taylor expansion, $\Lambda^{\mathrm{lr}}(u_{k,l})$ can be approximated by:

$$
\begin{aligned}
&\log\left(1-\frac{R}{2}+\frac{R}{2}\left(1+\frac{\Delta\,\mathrm{sign}(\Delta k-\mu)(k-\bar{k})}{b}\right)\right)\\
\approx&\log\left(1+\left(\frac{R\Delta\,\mathrm{sign}(\Delta k-\mu)(k-\bar{k})}{2b}\right)\right),\\
\approx&\frac{R\Delta}{2b}(k-\bar{k})\,\mathrm{sign}(\Delta k-\mu).
\end{aligned}
$$

Appendix C: LR Based on the Gaussian Model (WS)

Let X be a Gaussian random variable with expectation μ and variance σ^2. Define the quantized Gaussian random variable as follows $Y=\lfloor X/\Delta\rceil$, its pmf is given by $P_{\mu,\sigma}=\{p_{\mu,\sigma}[k]\}_{k=-\infty}^{\infty}$ with:

$$
p_{\mu,\sigma}[k]=\mathbb{P}[Y=k]=\int_{\Delta(k-1/2)}^{\Delta(k+1/2)}\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)dx.
$$

Assuming that the quantization step Δ is "small enough" compared to the variance $\Delta << \sigma$, it holds true that [6, 44]:

$$
p_{\mu,\sigma}[k]\approx\frac{\Delta}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(\Delta k-\mu)^2}{2\sigma^2}\right),
\tag{34}
$$

and

$$
p_{\mu,\sigma}[k]+p_{\mu,\sigma}[\bar{k}]\approx\frac{2\Delta}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(\Delta\frac{(k+\bar{k})}{2}-\mu)^2}{2\sigma^2}\right).
\tag{35}
$$

Let us rewrite the LR for the detection of JSteg (7) as follows

$$
\begin{aligned}
\Lambda^{\mathrm{lr}}(u_{k,l})&=\log\left(1-\frac{R}{2}+\frac{R}{2}\frac{p_{\mu,\sigma}[\bar{k}]}{p_{\mu,\sigma}[k]}\right),\\
&=\log\left(1-R+\frac{R}{2}\frac{p_{\mu,\sigma}[\bar{k}]+p_{\mu,\sigma}[k]}{p_{\mu,\sigma}[k]}\right).
\end{aligned}
\tag{36}
$$

Using the expressions (34) and (35) let us study the following ratio:

$$
\begin{aligned}
&\frac{p_{\mu,\sigma}[\bar{k}]+p_{\mu,\sigma}[k]}{p_{\mu,\sigma}[k]}=2\frac{\exp\left(-\frac{\Delta\frac{(k+\bar{k})}{2}-\mu)^2}{2\sigma^2}\right)}{\exp\left(-\frac{(\Delta k-\mu)^2}{2\sigma^2}\right)},\\
=&2\frac{\exp\left(-\frac{\left(\Delta k-\mu+\Delta/2(\bar{k}-k)\right)^2}{2\sigma^2}\right)}{\exp\left(-\frac{(\Delta k-\mu)^2}{2\sigma^2}\right)},\\
=&2\frac{\exp\left(-\frac{(\Delta k-\mu)^2}{2\sigma^2}\right)\exp\left(\frac{\Delta(\Delta k-\mu)(k-\bar{k})}{2\sigma^2}\right)\exp\left(-\frac{\Delta^2}{8\sigma^2}\right)}{\exp\left(-\frac{(\Delta k-\mu)^2}{2\sigma^2}\right)},\\
=&2\exp\left(\frac{\Delta(\Delta k-\mu)(k-\bar{k})}{2\sigma^2}\right)\exp\left(-\frac{\Delta^2}{8\sigma^2}\right).
\end{aligned}
\tag{37}
$$

Putting the expression (37) into the expression of the log-LR (36) immediately gives:

$$
\Lambda^{\mathrm{lr}}(u_{k,l})=\log\left(1+R\left(\exp\left(\frac{\Delta(\Delta k-\mu)(k-\bar{k})}{2\sigma^2}\right)\exp\left(-\frac{\Delta^2}{8\sigma^2}\right)-1\right)\right)
\tag{38}
$$

from which a Taylor expansion around $\Delta/\sigma=0$, this results from the assumption that $\Delta << \sigma$, and finally gives the well-known expression of the WS:

$$
\Lambda^{\mathrm{lr}}(u_{k,l})\approx\frac{R\Delta}{2\sigma^2}(k-\bar{k})(\Delta k-\mu)
\tag{39}
$$

References

1. Böhme, R.: Advanced Statistical Steganalysis. Springer, New York (2010)
2. Fridrich, J.: Steganography in Digital Media: Principles, Algorithms, and Applications. Cambridge University Press, Cambridge (2009)
3. Cox, I., Miller, M., Bloom, J., Fridrich, J., Kalker, T.: Digital Watermarking and Steganography. Morgan Kaufmann, Burlington (2007)
4. Fridrich, J., Kodovskỳ, J.: Steganalysis of lsb replacement using parity-aware features. In: Information Hiding, pp. 31–45 (2013). Springer

5. Zhang, T., Ping, X.: A fast and effective steganalytic technique against jsteg-like algorithms. In: Proceedings of the 2003 ACM Symposium on Applied Computing, pp. 307–311 (2003). ACM
6. Cogranne, R., Zitzmann, C., Fillatre, L., Retraint, F., Nikiforov, I., Cornu, P.: Statistical decision by using quantized observations. In: Information Theory Proceedings (ISIT), 2011 IEEE International Symposium On, pp. 1210–1214 (2011). IEEE
7. Cogranne, R., Zitzmann, C., Fillatre, L., Retraint, F., Nikiforov, I., Cornu, P.: A cover image model for reliable steganalysis. In: Information Hiding, pp. 178–192 (2011). Springer
8. 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 (2011). Springer
9. Fridrich, J., Goljan, M., Du, R.: Reliable detection of lsb steganography in color and grayscale images. In: Proceedings of the 2001 Workshop on Multimedia and Security: New Challenges, pp. 27–30 (2001). ACM
10. Dumitrescu, S., Wu, X., Wang, Z.: Detection of lsb steganography via sample pair analysis. Signal Processing, IEEE Transactions on **51**(7), 1995–2007 (2003)
11. Ker, A.D., Bas, P., Böhme, R., Cogranne, R., Craver, S., Filler, T., Fridrich, J., Pevnỳ, T.: Moving steganography and steganalysis from the laboratory into the real world. In: Proceedings of the First ACM Workshop on Information Hiding and Multimedia Security, pp. 45–58 (2013). ACM
12. Fridrich, J., Goljan, M.: On estimation of secret message length in lsb steganography in spatial domain. In: Electronic Imaging 2004, pp. 23–34 (2004). International Society for Optics and Photonics
13. 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 (2004)
14. Ker, A.D., Böhme, R.: Revisiting weighted stego-image steganalysis. In: Electronic Imaging 2008, pp. 681905–681905 (2008). International Society for Optics and Photonics
15. Böhme, R.: Weighted stego-image steganalysis for jpeg covers. In: Information Hiding, pp. 178–194 (2008). Springer
16. Cogranne, R., Zitzmann, C., Retraint, F., Nikiforov, I.V., Cornu, P., Fillatre, L.: A local adaptive model of natural images for almost optimal detection of hidden data. Signal Processing **100**, 169–185 (2014)
17. Lam, E.Y., Goodman, J.W.: A mathematical analysis of the dct coefficient distributions for images. Image Processing, IEEE Transactions on **9**(10), 1661–1666 (2000)
18. Muller, F.: Distribution shape of two-dimensional dct coefficients of natural images. Electronics Letters **29**(22), 1935–1936 (1993)
19. Chang, J.-H., Shin, J.W., Kim, N.S., Mitra, S.K.: Image probability distribution based on generalized gamma function. Signal Processing Letters, IEEE **12**(4), 325–328 (2005)
20. Sallee, P.: Model-based methods for steganography and steganalysis. International Journal of Image and graphics **5**(01), 167–189 (2005)
21. Böhme, R., Westfeld, A.: Breaking cauchy model-based jpeg steganography with first order statistics, 125–140 (2004)
22. Zitzmann, C., Cogranne, R., Fillatre, L., Nikiforov, I., Retraint, F., Cornu, P.: Hidden information detection based on quantized laplacian distribution. In: Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference On, pp. 1793–1796 (2012). IEEE
23. Kodovsky, J., Fridrich, J.: Quantitative structural steganalysis of jsteg. Information Forensics and Security, IEEE Transactions on **5**(4), 681–693 (2010)
24. Lee, K., Westfeld, A., Lee, S.: Category attack for lsb steganalysis of jpeg images. In: Digital Watermarking, pp. 35–48. Springer, New York (2006)
25. Lyu, S., Farid, H.: Steganalysis using higher-order image statistics. Information Forensics and Security, IEEE Transactions on **1**(1), 111–119 (2006)
26. Pevny, T., Fridrich, J.: Multiclass detector of current steganographic methods for jpeg format. Information Forensics and Security, IEEE Transactions on **3**(4), 635–650 (2008)
27. Bas, P., Filler, T., Pevný, T.: Break our steganographic system — the ins and outs of organizing boss. In: Filler, T. (ed.) Information Hiding, 13th International Workshop (2011)
28. Denemark, T., Sedighi, V., Holub, V., Cogranne, R., Fridrich, J.: Selection-channel-aware rich model for steganalysis of digital images
29. Tang, W., Li, H., Luo, W., Huang, J.: Adaptive steganalysis against wow embedding algorithm. In: Proceedings of the 2nd ACM Workshop on Information Hiding and Multimedia Security, pp. 91–96 (2014). ACM
30. Qiao, T., Ziitmann, C., Cogranne, R., Retraint, F.: Detection of jsteg algorithm using hypothesis testing theory and a statistical model with nuisance parameters. In: Proceedings of the 2nd ACM Workshop on Information Hiding and Multimedia Security, pp. 3–13 (2014). ACM
31. Qiao, T., Zitzmann, C., Cogranne, R., Retraint, F.: Statistical detection of jsteg steganography using hypothesis testing theory. In: IEEE International Conference on Image Processing (ICIP), pp. 1–5 (2014)
32. Nakamura, J.: Image Sensors and Signal Processing for Digital Still Cameras. CRC Press, Boca Raton (2005)
33. Pennebaker, W.B., Mitchell, J.L.: JPEG: Still Image Data Compression Standard. Springer, Germany (1993)
34. Upham, D.: Jsteg steganographic algorithm. Available on the internet http://www.filewatcher.com/m/jpeg-jsteg-v4.diff.gz.8878-0.html (1999)
35. Lehmann, E.L., Romano, J.P.: Testing Statistical Hypotheses. Springer, Germany (2006)
36. Thai, T.H., Cogranne, R., Retraint, F.: Steganalysis of jsteg algorithm based on a novel statistical model of quantized dct coefficients. In: ICIP, pp. 4427–4431 (2013)
37. Thai, T., Cogranne, R., Retraint, F.: Statistical model of quantized dct coefficients: Application in the steganalysis of jsteg algorithm. IEEE transactions on image processing: a publication of the IEEE Signal Processing Society **23**(5), 1980 (2014)
38. Cogranne, R., Retraint, F.: An asymptotically uniformly most powerful test for lsb matching detection, 464–476 (2013)
39. Thai, T.H., Retraint, F., Cogranne, R.: Statistical model of natural images. In: Image Processing (ICIP), 2012 19th IEEE International Conference On, pp. 2525–2528 (2012). IEEE
40. Lebrun, M.: An analysis and implementation of the bm3d image denoising method. Image Processing On Line (2012)
41. Lebrun, M., Leclaire, A.: An implementation and detailed analysis of the k-svd image denoising algorithm. Image Processing On Line (2012)
42. Buades, A., Coll, B., Morel, J.-M.: A non-local algorithm for image denoising. In: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference On, vol. 2, pp. 60–65 (2005). IEEE
43. Lukas, J., Fridrich, J., Goljan, M.: Digital camera identification from sensor pattern noise. Information Forensics and Security, IEEE Transactions on **1**(2), 205–214 (2006)
44. Cogranne, R., Retraint, F., Zitzmann, C., Nikiforov, I., Fillatre, L., Cornu, P.: Hidden information detection using decision theory and quantized samples: Methodology, difficulties and results. Digital Signal Processing **24**, 144–161 (2014)