

# Statistical detection of JPEG traces in digital images in uncompressed formats

Cecilia Pasquini, Giulia Boato, *Member, IEEE*, and Fernando Pérez-González, *Fellow, IEEE*

**Abstract**—Intrinsic statistical properties of natural uncompressed images are used in image forensics for detecting traces of previous processing operations. In this paper, we propose novel forensic detectors of JPEG compression traces in images stored in uncompressed formats, based on a theoretical analysis of Benford–Fourier coefficients computed on the  $8 \times 8$  block-DCT domain. In fact, the distribution of such coefficients is derived theoretically both under the hypotheses of no compression and previous compression with a certain quality factor, allowing for the computation of the respective likelihood functions. Then, two classification tests based on different statistics are proposed, both relying on a discriminative threshold that can be determined without the need of any training phase. The statistical analysis is based on the only assumptions of Generalized Gaussian distribution of DCT coefficients and independence among DCT frequencies, thus resulting in robust detectors applying to any uncompressed image. In fact, experiments on different datasets show that the proposed models are suitable for images of different sizes and source cameras, thus overcoming dataset-dependency issues that typically affect state-of-art techniques.

**Index Terms**—Benford–Fourier coefficients, Benford’s law, image forensics, JPEG compression.

## I. INTRODUCTION

IN the last decade, the field of digital image forensics has undergone a constant expansion, drawing a growing attention of the research community. In fact, the increasing need of authentication techniques for digital images without *a priori* information (like the presence of a watermark) has brought researchers to develop a number of forensic approaches designed to work passively and facing different forensic issues, like forgery detection, source identification, and computer generated versus natural content discrimination [1] [2] [3]. One of the most widely investigated problems is the detection of previous operations and the estimation of the processing history, which might reveal the non-pristine condition of the subject image and allow for the localization of forged areas. Given the wide variety of potential processing operations and the corresponding statistical traces, a high number of forensic techniques have been designed, often tailored to specific traces and experimental settings. Several valuable solutions have

been proposed for different forensic problems, yielding good performance both in synthetic and more realistic forensic scenarios. However, a common issue to many methods is the fact that a solid theoretic framework describing the statistical behavior of the quantities involved is not available. Thus, although they achieve excellent results in certain experimental settings, the absence of a generalized model might result in non-controllable performance when the test setting is modified since the parameters of the methods change as well. This generally happens, for instance, when a certain approach implies the need of machine learning techniques: although they represent extremely useful tools, they usually require extensive training phases and suffer from typical automatic learning issues (as overfitting or dataset-dependency), which might have a strong impact on the applicability of multimedia forensic techniques in different settings. This is especially true when data depicting sensitive content are analyzed and forensic analysis reliability is an essential requirement, as it commonly happens in real-world cases [4] and even more so when they involve courts of law.

In this paper, we tackle from a theoretical perspective the problem of detecting the traces of a previous JPEG compression in images that are stored in uncompressed formats. Such issue appears when the forensic analysis is performed on images supposedly taken by a device set to provide uncompressed images (like professional or semi-professional cameras) or, in general, in every situation where the subject image is supposed to be never compressed, and the presence of JPEG compression traces would suggest that the image has been taken from a different camera or it has been already processed by someone. Indeed, although the JPEG standard represents the most used format for digital images, the need for analyzing uncompressed formats arises, for instance, when professional photographic images are involved, for which it is getting a common practice to provide the original raw image and employ it in the forensic analysis [5].

In this framework, the proposed method relies on the theoretical analysis of the Benford–Fourier (BF) coefficients [6] computed from the DCT coefficients of the image, for which a statistical model is derived both under the hypotheses of no previous compression (i.e., the image is pristine) and previous compression with a certain quantization table (i.e., the image has been JPEG compressed). This allows us to define a hypothesis testing framework where the null hypothesis is the pristine condition of the image, and the alternative hypothesis is represented by a previous compression. The work [7] contains a preliminary and partial version of the methodology, which is here extended to the case multiple

C. Pasquini was with the Department of Information Engineering and Computer Science, University of Trento, Trento 38123, Italy. She is now with the Institute für Informatik, Universität Innsbruck, Innsbruck 6020, Austria, and also with the Department of Information Systems, University of Münster, Münster 48149, Germany (e-mail: cecilia.pasquini@uibk.ac.at).

G. Boato is with the Department of Information Engineering and Computer Science, University of Trento, Trento 38123, Italy (e-mail: giulia.boato@unitn.it).

F. Pérez-González is with the Department of Signal Theory and Communications, University of Vigo, Vigo 36310, Spain (e-mail: fperez@gts.uvigo.es).

Manuscript received xxx xx, xxxx; revised xxx xx, xxxx.

DCT frequencies and encompasses also a novel model for the alternative hypotheses. This results in two novel tests based on different statistical schemes, namely the  $\lambda$ -test and the  $\log L_0$ -test, with the aim of discriminating images that have never been compressed from images previously compressed.

An interesting peculiarity of the proposed methods is that the statistical description of the BF coefficients, derived analytically, explicitly depends on the number of DCT coefficients considered, i.e., it is related to the size of the subject image. Moreover, all the statistical parameters involved in the model are estimated directly from the data without relying on any predetermined dataset. This results in two size-adaptive JPEG compression detectors, which do not require any training phase. Experimental results on several datasets and JPEG compression parameters show the benefits of this approach with respect to state-of-the-art methods.

The paper is structured as follows: in Section II, the literature on the detection of JPEG traces in uncompressed format images is revisited; the BF coefficients are introduced in Section III; the novel statistical analysis of such coefficients is presented in Section IV and in Section V the proposed detection algorithm is described; the experimental tests performed are reported in Section VI, while conclusions are presented in Section VII.

## II. RELATED WORK

The detection of JPEG compression traces in digital images that are stored in uncompressed formats is a known problem in multimedia forensic research. In principle, a previous compression could be identified by exhaustively recompressing the test image with all the possible quality factors and looking for the minimal distance between the test image and the recompressed versions. However, such an approach is extremely time consuming for bigger images and highly dependent on the image content, leading to poor detection rates. As a matter of fact, researchers have developed several alternative techniques, each of them relying on different statistics and presenting pros and cons.

One of the first approaches was proposed in [8]: there, the blocking artifacts left by a JPEG compression in the pixel domain are exploited, and a detector based on inter- and intra-block pixel differences is designed. Such values are combined in a final statistic  $K$ , expressing the strength of blocking artifacts, and images presenting a value of  $K$  higher than a certain threshold are classified as compressed. In the same paper, a procedure based on ML estimation of the used quantization table is proposed. An improved version is presented in [9], where the joint detection of both the quantization table and the used color space transformation is achieved.

As the quantization of the  $8 \times 8$ -block DCT coefficients represents the core of the JPEG compression procedure and leaves characteristic footprints, several methods for JPEG images focus on the analysis of the DCT coefficients for extracting information regarding the compression history. In [10], the distribution of DCT coefficients after quantization and reprojecting on the pixel domain is studied: in particular,

the authors observe how the DCT coefficients behave differently around 0 when the image is pristine or previously compressed. Such different behaviors are captured in a 1D feature, discriminating between original and compressed images; in the latter case, a simple procedure is proposed for estimating the quantization steps. A similar rationale is exploited in [11], where the authors suggest to use the variance of the forward quantization noise (error in quantizing the DCT coefficients) as discriminative threshold. On the other hand, the work presented in [12] studies the first order statistics of the factors of the AC DCT coefficients, i.e., the set of numbers that divide evenly the coefficients: after the quantization the histogram of the factor set is altered, showing sharper peaks. Thus, the maximum difference between adjacent bins is considered as discriminative threshold.

Such methods are characterized by a low complexity and good performance, also in case of small images; on the other hand, the used statistics present a quite different behaviour when varying the size and source camera of the image and therefore the performance is strongly dependent on the initial set of images used for determining the threshold.

Another statistic that has been explored in image forensics is the distribution of the First Significant Digits (FSD) of the DCT coefficients. Indeed, when the DCT coefficients are quantized, their FSDs change together with their distribution. In particular, for uncompressed images we have that the FSDs follow a logarithmic distribution, known as Benford's law, which is perturbed when a quantization occurs. Driven by this observation, the authors in [13] proposed a JPEG compression detector based on an SVM classifier which uses as features the empirical frequencies of the nine FSDs on all the DCT coefficients in the image. The method achieves good results on the considered dataset and requires a relatively low computational complexity; however, it does not provide an estimate of the quality factor or quantization table used, since no theoretical model for the FSD distribution is available, and the results are strongly dependent on the dataset.

Recently, a first approach based of Benford–Fourier coefficients has been proposed in [6] and a preliminary version of this work is available in [7]; both of them will be described in detail in the next sections.

## III. BENFORD–FOURIER COEFFICIENTS

In this work, we exploit Benford–Fourier coefficients to characterize uncompressed images and identify potential previous processing operations, in particular JPEG compression. Such coefficients have been originally introduced in [14] and have a precise mathematical meaning which makes them extremely suitable for the considered forensic problem.

For the sake of clarity, in the following we will indicate univariate real or complex random variables with capital letters, whose realizations will be represented by the corresponding lower case letters. Moreover, we will denote the real and imaginary parts of a complex number  $a$  as  $\Re(a)$  and  $\Im(a)$ , respectively.

Then, let  $X$  be a random variable representing the non-zero DCT coefficients and  $f_X$  its probability density function;

TABLE I  
MAGNITUDE OF THE BENFORD–FOURIER COEFFICIENT  $a_n$  FOR DIFFERENT VALUES OF THE SHAPING FACTOR  $\nu$  OF THE GG DISTRIBUTION AND  $n$ .

$n$	1	2	3	4	5	6	7	8
$\nu = 0.5$	$6.1 \cdot 10^{-3}$	$3.2 \cdot 10^{-6}$	$1.1 \cdot 10^{-9}$	$3.3 \cdot 10^{-13}$	$9.0 \cdot 10^{-17}$	$2.2 \cdot 10^{-20}$	$5.5 \cdot 10^{-24}$	$1.3 \cdot 10^{-27}$
$\nu = 0.75$	$2.7 \cdot 10^{-2}$	$1.6 \cdot 10^{-4}$	$7.4 \cdot 10^{-7}$	$3.1 \cdot 10^{-9}$	$1.2 \cdot 10^{-11}$	$4.8 \cdot 10^{-14}$	$1.8 \cdot 10^{-16}$	$6.7 \cdot 10^{-19}$
$\nu = 1$	$5.6 \cdot 10^{-2}$	$1.1 \cdot 10^{-3}$	$1.8 \cdot 10^{-5}$	$2.9 \cdot 10^{-7}$	$4.6 \cdot 10^{-9}$	$6.9 \cdot 10^{-11}$	$1.0 \cdot 10^{-12}$	$1.5 \cdot 10^{-14}$
$\nu = 1.25$	$8.8 \cdot 10^{-2}$	$3.5 \cdot 10^{-3}$	$1.2 \cdot 10^{-4}$	$4.5 \cdot 10^{-6}$	$1.5 \cdot 10^{-7}$	$5.4 \cdot 10^{-9}$	$1.8 \cdot 10^{-10}$	$6.2 \cdot 10^{-12}$

we suppose  $f_X$  is symmetric with respect to 0. Then, we define  $Z$  as the random variable representing the absolute value of nonzero DCT coefficients. In other words,  $Z$  models the behavior of  $|X|$  in  $\mathbb{R}_0^+$  (i.e., discarding the value 0 as possible outcome), and its pdf is given by

$$f_Z(z) = 2 \cdot f_X(z) \quad \forall z \in \mathbb{R}_0^+,$$

as  $f_X$  is symmetric.

Then, we define the random variables

$$Z' \doteq \log_{10} Z$$

$$\tilde{Z} \doteq \log_{10} Z \pmod{1},$$

representing the absolute valued positive DCT coefficients in the logarithmic and modular logarithmic domains, respectively. The r.v.  $\tilde{Z}$  is particularly relevant because of its relationship with the pdf of the FSD of  $X$  [14] and it has been exploited in JPEG counter-forensic techniques [15] [16].

Now, the *Benford–Fourier (BF) coefficients* in  $n \in \mathbb{N}$  are defined as the Fourier transform of  $f_{Z'}(z')$  evaluated at  $2\pi n$ , i.e.,

$$a_n = \int_{-\infty}^{+\infty} f_{Z'}(z') e^{-j2\pi n z'} dz' = \int_{-\infty}^{+\infty} f_Z(z) e^{-j2\pi n \log_{10} z} dz. \quad (1)$$

Such coefficients turn out to be particularly suitable for characterizing the DCT coefficient behavior since they have a key role in the statistical description of  $\tilde{Z}$ : in fact, in [14] it has been shown that, for a generic continuous r.v.  $Z$ , we have

$$f_{\tilde{Z}}(\tilde{z}) = 1 + 2 \sum_{n=1}^{+\infty} |a_n| \cos(2\pi n \tilde{z} + \phi_n), \quad \tilde{z} \in [0, 1). \quad (2)$$

where  $a_n = |a_n| e^{j\phi_n}$ .

Moreover, it has been shown that the DCT coefficients of uncompressed images can be modeled by a Generalized Gaussian r.v. with a shaping factor generally ranging from 0.5 to 1.2 [17] and we adopt such model throughout the following derivations. We choose not to study modeling errors that could potentially occur, since other effects introduced in the compression process (like quantization in the pixel domain) will prevail on them. In [14] the authors show that if  $X$  is a Generalized Gaussian (GG) r.v. with standard deviation  $\sigma$  and shaping factor  $\nu$ , i.e.,

$$f_X(x) = A e^{-|\beta x|^\nu}, \quad x \in \mathbb{R},$$

$$\beta = \frac{1}{\sigma} \sqrt{\frac{\Gamma(3/\nu)}{\Gamma(1/\nu)}}, \quad A = \frac{\beta \nu}{2\Gamma(1/\nu)},$$

the theoretical expression of  $a_n$ ,  $n \in \mathbb{N}$ , and its magnitude can be derived as functions of the GG parameters:

$$a_n = \int_{-\infty}^{+\infty} f_{Z'}(z') e^{-j2\pi n z'} dz' = \frac{2A}{\beta \nu} e^{j \frac{2\pi n \log \beta}{\log 10}} \Gamma\left(\frac{-j2\pi n + \log 10}{\nu \log 10}\right),$$

$$|a_n|^2 = \prod_{k=0}^{\infty} \left[ 1 + \frac{(2\pi n)^2}{\log^2 10 (\nu k + 1)^2} \right]^{-1}, \quad (3)$$

where  $\log$  denotes the natural logarithm.

As seen in (3), the magnitude of the coefficients increases with  $\nu$  and does not depend on the variance of the GG. In Table I, we report the values of  $|a_n|$  computed as in (3) (where  $k$  ranges from 0 to  $10^5$ ) for different values of  $n$  and  $\nu$ . We can notice that, in particular, when  $n \geq 3$ , the  $|a_n|$  are always lower than  $10^{-4}$ .

This represents a useful information in JPEG image forensics and suggests that the behavior of the BF coefficients can be used to characterize uncompressed images. Indeed, a first approach in this direction was proposed in [6], where the BF coefficients from the DCT coefficients of the whole image are estimated by computing the FFT of the empirical distribution of  $\tilde{Z}$ . Then, the first five coefficients (i.e.,  $n = 1, \dots, 5$ ) are used as feature to train an SVM discriminating between natural uncompressed images and images that underwent a JPEG compression, obtaining promising results.

For further details on the role of Benford's law and Benford–Fourier coefficients in images, a comprehensive treatment can be found in [6].

#### IV. STATISTICAL ANALYSIS OF BENFORD–FOURIER COEFFICIENTS

In this work, we aim at discriminating compressed images saved in uncompressed format from images that underwent a JPEG compression. Accordingly, for each DCT frequency we want to quantify the probabilities that the DCT coefficients have never been quantized or that they have been previously quantized with a generic step  $q$ . To this aim, we consider the BF coefficients at a fixed DCT frequency and develop a statistical model for each of these two cases. Such models will then be exploited in the final test, where the hypotheses of no compression and compression with a quality factor among a predetermined pool are considered. Whereas the hypothesis testing scheme will be described in detail in Section V, in the following we present the statistical models derived for the BF coefficient of a single DCT frequency. A preliminary version of the model in Section IV-A was proposed in [7],

while Section IV-B contains a novel analysis of the quantized DCT coefficients.

### A. Uncompressed image model

In order to use BF coefficients for analyzing an image, we need a numerical procedure to estimate them given the subject image.

By looking at (1), we can notice that  $a_n$  is the expected value of the complex random variable  $g_n(Z) = e^{-j2\pi n \log_{10} Z}$ , whose values lie on the unit circle<sup>1</sup>. Thus, as it is usually done in statistics, we can obtain an estimate of  $a_n = E\{g_n(Z)\}$  by considering the sample mean of  $g_n(Z)$  provided by the DCT coefficients of the image through the different  $8 \times 8$  blocks. Thus, if we denote as  $z^m$  the value of the DCT coefficient at the chosen frequency in the  $m$ -th block, we can consider as estimator of  $a_n$  the expression

$$\hat{a}_n \doteq \frac{\sum_{m=1}^M e^{-j2\pi n \log_{10} z^m}}{M}, \quad m = 1, \dots, M, \quad (4)$$

where  $M$  is the total number of  $8 \times 8$  blocks in the image.

In other words, we can see  $\hat{a}_n$  as realization of the r.v.

$$\hat{A}_n \doteq \frac{\sum_{m=1}^M e^{-j2\pi n \log_{10} Z^m}}{M}, \quad m = 1, \dots, M, \quad (5)$$

where  $Z^m$  is the r.v. representing the DCT coefficient at the chosen frequency in the  $m$ -th block.

Although the sample mean is a minimum variance unbiased estimator of the expected value (i.e.,  $E\{\hat{A}_n\} = a_n$ ), we should take into account the fact that the actual accuracy of  $\hat{a}_n$  in the estimation of  $a_n$  depends on the size of the considered sample. For this reason, we are interested in studying the distribution of  $\hat{A}_n$  as a function of the number of samples  $M$ .

To this end, we can observe that  $\hat{A}_n$  is a sum of  $M$  independent and identically distributed random variables  $g_n(Z^m)$ . Then, by applying the Central Limit Theorem (CLT) to the real and imaginary parts of  $\hat{A}_n$ , we have that their distribution is asymptotically Gaussian with expected values  $\Re(a_n)$  and  $\Im(a_n)$ , respectively [18]. In other words,

$$\hat{A}_n \approx a_n + W_0,$$

where  $W_0$  is a zero-mean complex normal random variable.

A necessary and sufficient condition for  $W_0$  to be circularly symmetric (i.e., with real and imaginary parts independent and identically distributed [18]) is that  $E\{W_0^2\} = 0$ . Starting from the definition of  $\hat{A}_n$ , it is easy to prove that

$$E\{W_0^2\} \approx E\{(\hat{A}_n - a_n)^2\} = \frac{1}{M}(a_{2n} - a_n^2). \quad (6)$$

Hence,  $|E\{W_0^2\}| \leq (|a_{2n}| + |a_n^2|)/M$  and, by looking at Table I, we can conclude that the value of (6) will be very close to 0 (for instance, when  $\nu = 1$  and  $n = 3$  its order of magnitude is  $10^{-11}$ ). Therefore,  $\hat{A}_n$  is approximately a circular bivariate normal r.v. with non-zero mean.

<sup>1</sup>It is worth noticing that BF coefficients are defined for integers  $n \in \mathbb{N}$  but such definition could be easily extended to the entire real line by replacing  $2\pi n$  with a real value  $\omega$ , and the following analysis would hold identically also in this more general case.

It is well known that the r.v.  $R \doteq |\hat{A}_n|$  approximately follows a Rice distribution with mean parameter  $|a_n|$  and scale parameter  $s$ , where  $s$  is the standard deviation of both its real and imaginary parts [19]. Similarly as before, we can now obtain  $s^2$  by exploiting the fact that for a Rice distribution

$$s^2 = \frac{E\{|\hat{A}_n|^2\} - |a_n|^2}{2} = \frac{1}{2M}(1 - |a_n|^2).$$

As we observed,  $|a_n|$  is lower than  $10^{-4}$  when  $n \geq 3$  and we can reasonably assume  $|a_n| \approx 0$ , thus considering the special case of Rice distribution with mean parameter 0, i.e., the Rayleigh distribution with scale parameter  $s = 1/\sqrt{2M}$ . According to this, we can define  $p(\hat{a}_n|NQ)$  ( $NQ$  means “never quantized”) as the probability density function of obtaining a BF coefficient  $\hat{a}_n$  under the hypothesis of no previous quantization and compute it as follows:

$$p(\hat{a}_n|NQ) = 2M|\hat{a}_n|e^{-M|\hat{a}_n|^2}, \quad (7)$$

where the expression on the right is the Rayleigh pdf with  $s = 1/\sqrt{2M}$ . By considering its properties, we have that  $|\hat{a}_n|$  is in any case an overestimate of  $|a_n| = 0$ , where its mean is given by  $\frac{1}{\sqrt{M}} \cdot \frac{\sqrt{\pi}}{2}$  (the expected accuracy increases linearly with  $\sqrt{M}$ ) and its variance is given by  $\frac{1}{M} \cdot \frac{4-\pi}{4}$  (the expected accuracy variance decreases linearly with  $M$ ).

An example of the model is showed in Fig. 1.

### B. Compressed image model

When computing the DCT from an image stored in uncompressed format that was previously compressed, the DCT coefficients at a certain frequency have a distribution like in Fig. 2a. The error affecting the histogram is due to the quantization in the pixel domain after the block-DCT quantization and the rounding/truncation errors in the DCT computation, and has been modeled in the literature as a Gaussian r.v. [20].

We propose here an alternative statistical description whose accuracy has been assessed by extensive numerical tests. We can restrict our analysis to a single quantization interval and consider the DCT coefficients contained within it. Without any loss of generality, we consider the interval  $I_q \doteq [q - q/2, q + q/2[$  ( $q$  is the quantization step) and we denote with  $Z_q$  the r.v. representing the DCT coefficients falling in  $I_q$ . Then, we can approximate its distribution with a Laplacian truncated outside the quantization interval, as in Fig. 2b. Then, the pdf of  $Z_q$  is given by

$$f_{Z_q}(z) = \frac{\mathcal{L}(z; q, \sigma)}{N_{\sigma, q}} \cdot \mathbb{1}_{I_q}(z), \quad (8)$$

where  $\mathcal{L}(\cdot; q, \sigma)$  is a Laplacian pdf with mean  $q$  and standard deviation  $\sigma$  (which is unknown and needs to be estimated),  $N_{\sigma, q}$  is the integral of  $\mathcal{L}(z; q, \sigma)$  over  $I_q$  (so that expression (8) is a pdf), and  $\mathbb{1}_I(\cdot)$  is the indicator function of  $I_q$

$$\mathbb{1}_{I_q}(z) \doteq \begin{cases} 1 & z \in I_q \\ 0 & z \notin I_q. \end{cases}$$

Starting from this hypothesis, we can define  $a_{n,q}$  as the Benford-Fourier coefficients of  $Z_q$ , and derive their theoretical

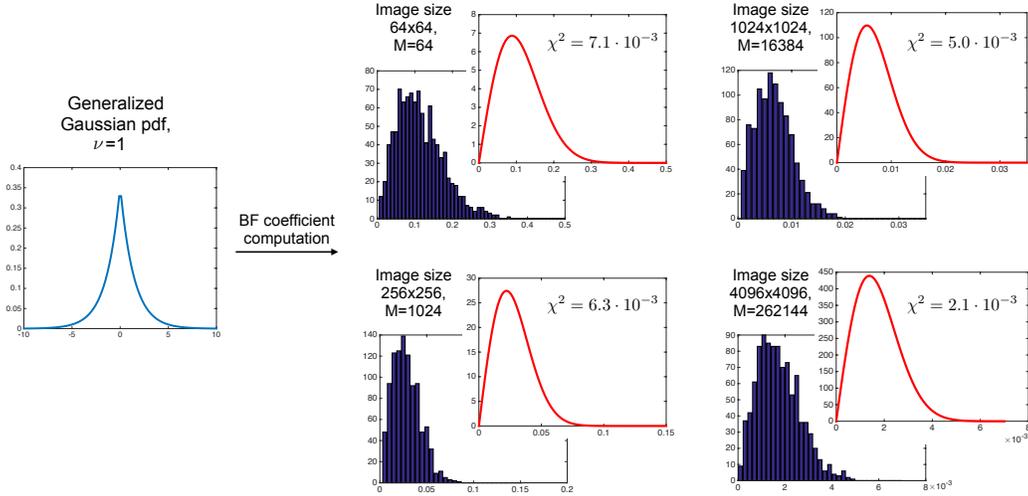


Fig. 1. The figure depicts the behavior of BF coefficients when varying the number  $M$  of samples used in the sum (4). For  $M = 64, 1024, 16384, 262144$ , we generated 1000 sets of  $M$  elements distributed according to a zero-mean Generalized Gaussian pdf with  $\nu = 1$  (which is common for DCT coefficients of natural images) and varying standard deviation. Then, the estimates of BF coefficient magnitude  $|\hat{a}_n|$  (with  $n = 4$ ) have been computed on each set and their histograms for the different values of  $M$  are plotted (blue bars). The red curves are the Rayleigh pdfs (7) determined by the corresponding value of  $M$  in each case. The goodness of fit of the proposed model is confirmed by the values of the  $\chi^2$  statistics (computed on 10 equally spaced bins from 0 to the highest magnitude value obtained in each case).

value as follows:

$$\begin{aligned}
 a_{n,q} &\doteq \int_{-\infty}^{+\infty} f_{Z_q}(z) e^{-j2\pi n \log_{10} z} dz \\
 &= \frac{1}{N_{\sigma,q}} \int_{I_q} \frac{1}{\sigma\sqrt{2}} e^{-\frac{\sqrt{2}}{\sigma}|z-q|} e^{-j2\pi n \log_{10} z} dz \\
 &= \frac{1}{N_{\sigma,q}\sigma\sqrt{2}} \left( e^{-\frac{\sqrt{2}}{\sigma}kq} \int_{q-q/2}^q e^{\frac{\sqrt{2}}{\sigma}z} z^{-j\frac{2\pi n}{\log_{10}}} dz + \right. \\
 &\quad \left. + e^{\frac{\sqrt{2}}{\sigma}q} \int_q^{q+q/2} e^{-\frac{\sqrt{2}}{\sigma}z} z^{-j\frac{2\pi n}{\log_{10}}} dz \right). \quad (9)
 \end{aligned}$$

In other words, assuming a Laplacian distribution of  $Z_q$  and

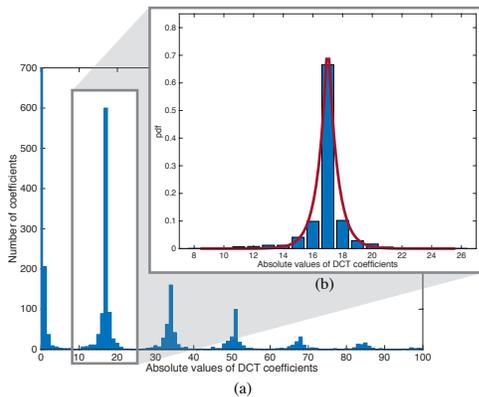


Fig. 2. In panel (a), histogram of DCT coefficients at a single frequency after quantization with  $q = 17$ . In panel (b), the coefficients corresponding to the r.v.  $Z_{17}$  are reported and the red curve is the Laplacian pdf estimated from the histogram.

given an estimate of  $\sigma$ , we can obtain the theoretical value of  $a_{n,q}$  from the previous expression in (9) by numerically computing the integrals.

Now, we can adopt the same approach as the uncompressed case: consider the sample mean

$$\hat{a}_{n,q} = \frac{\sum_{m=1}^{M_q} e^{-j2\pi n \log_{10} z_q^m}}{M_q}, \quad m = 1, \dots, M_q, \quad (10)$$

as estimator of  $a_{n,q}$ , which is the realization of the r.v.

$$\hat{A}_{n,q} = \frac{\sum_{m=1}^{M_q} e^{-j2\pi n \log_{10} Z_q^m}}{M_q}, \quad m = 1, \dots, M_q, \quad (11)$$

where  $M_q$  is the number of DCT coefficients falling in the interval  $I_q$  at the chosen frequency. Then, we need to study its distribution in order to obtain an expression of  $p(\hat{a}_{n,q}|q)$ , i.e., the probability of obtaining  $\hat{a}_{n,q}$  under the hypothesis that the DCT coefficients at the chosen frequency underwent a quantization with step  $q$ .

We can partially exploit the logical steps of the uncompressed case reported in Section IV-A. Indeed, exactly in the same way, we can conclude that

$$\hat{A}_{n,q} \approx a_{n,q} + W_{0,q},$$

where  $W_{0,q}$  is a complex zero-mean Gaussian random variable. In addition

$$E\{W_{0,q}^2\} = \frac{1}{M_q} (a_{2n,q} - (a_{n,q})^2).$$

Differently from the  $NQ$  case, we have no clue on the magnitudes of  $a_{2n,q}$ ,  $a_{n,q}$ , thus we cannot claim that  $E\{W_{0,q}^2\} \approx 0$  and  $W_{0,q}$  is circularly symmetric.

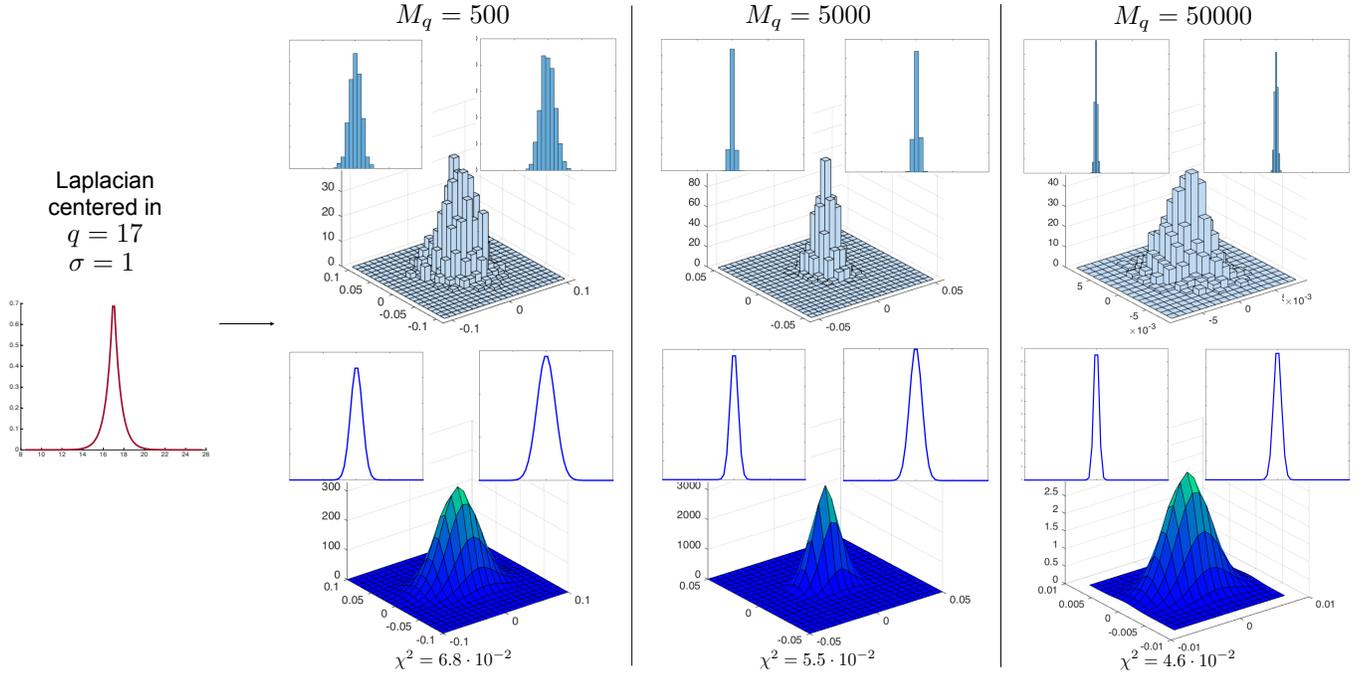


Fig. 3. The figure depicts the behavior of BF coefficients in case of compression when varying the number of samples belonging to the interval  $I_q$  considered. We studied the case of  $q = 17$  and, for  $M_q = 500, 5000, 50000$ , we generated 1000 sets of  $M_q$  elements distributed according to a Laplacian pdf with mean 17 and standard deviation  $\sigma = 1$ . Then, the estimate of the complex BF coefficient  $\hat{a}_{n,17}$  (with  $n = 4$ ) has been computed on each set and their histograms (for the complex values, the real and imaginary parts) for the different values of  $M_q$  are plotted (pale blue bars). The blue curves beside are the theoretical pdf for the complex r.v. and the marginal pdfs of the real and imaginary parts, determined as in Section IV-B. The match between the histograms obtained and the pdfs derived theoretically is confirmed by the value of the  $\chi^2$  statistics, computed on 25 two-dimensional bins (5 along the real part and 5 along the imaginary part).

Because of that, we consider the distribution of  $\hat{A}_{n,q}$  in the complex plane and we study the real and imaginary parts of  $W_{0,q}$ . For the sake of simplicity, we will denote them as  $W_r$  and  $W_i$ , respectively, (i.e., dropping the dependence on  $n$  and  $q$ ) and treat their joint pdf as a zero-mean real bivariate Gaussian<sup>2</sup>.

The analysis is slightly harder than the  $NQ$  case, since here we need to determine the three parameters of a real bivariate Gaussian: the variances of  $W_r$  and  $W_i$  ( $\sigma_{W_r}^2$  and  $\sigma_{W_i}^2$ , respectively) and the covariance between  $W_r$  and  $W_i$ ,  $Cov(W_r, W_i)$ .

All of them have been theoretically derived in order to obtain closed form expressions and are reported in the following:

- $\sigma_{W_r}^2, \sigma_{W_i}^2$

We have that  $\sigma_{W_r}^2$  and  $\sigma_{W_i}^2$  are given by the variances of the r.v.'s  $\Re(e^{-j2\pi n \log_{10} Z_q})$  and  $\Im(e^{-j2\pi n \log_{10} Z_q})$ , respectively, divided by the number of summands in the sample mean (10). The exact expressions of  $\sigma_{W_r}^2$  and  $\sigma_{W_i}^2$  are derived in the Appendix A in formula (23) and depend on  $n$ ,  $\sigma$  and  $q$ .

Note that an estimate of  $\sigma$  (the parameter of the Laplacian) is necessary and can be obtained from the data by means of an unbiased sample variance.

- $Cov(W_r, W_i)$

<sup>2</sup>In particular, we treat  $W_{0,q}$  as a real bivariate r.v. instead of a complex normal r.v., i.e., in terms of variance and covariance of the two single parts instead of complex covariance and pseudo-covariance (as it is usually done when dealing with complex r.v.'s).

We have that

$$\begin{aligned} E\{W_{0,q}\} &= E\{W_r^2\} - E\{W_i^2\} + 2jE\{W_r W_i\} \\ &= \frac{a_{2n,q} - (a_{n,q})^2}{M_q} \end{aligned}$$

and  $Cov(W_r, W_i) = E\{W_r W_i\}$ . Then, we can obtain

$$Cov(W_r, W_i) = \frac{\Im\left(\frac{1}{M_q}(a_{2n,q} - (a_{n,q})^2)\right)}{2}. \quad (12)$$

Fig. 3 depicts an example of the pdf's obtained by fixing a Laplacian distribution and generating sample vectors with varying length  $M_q$ . It can be noticed that the derived statistical models fit the data very accurately.

Finally, we can summarize the necessary steps to obtain  $p(\hat{a}_{n,q}|q)$  as follows:

- identify the set of DCT coefficients falling in  $I_q$ ,
- estimate the parameter  $\sigma$  of the Laplacian distribution by means of an unbiased sample variance,
- compute the theoretical value of  $a_{n,q}$  by means of (9) (in this phase numerical integration will be used),
- compute  $\sigma_{W_r}^2$ ,  $\sigma_{W_i}^2$  as in (23), and  $Cov(W_r, W_i)$  by means of (12),
- compute the estimate  $\hat{a}_{n,q}$  from the DCT coefficients as in (10)
- compute the probability of obtaining  $\hat{a}_{n,q}$  under the hypothesis of quantization with step  $q$  as follows

$$p(\hat{a}_{n,q}|q) = \mathcal{N}_2(\mu_{n,q}, \Sigma_{n,q}) \quad (13)$$

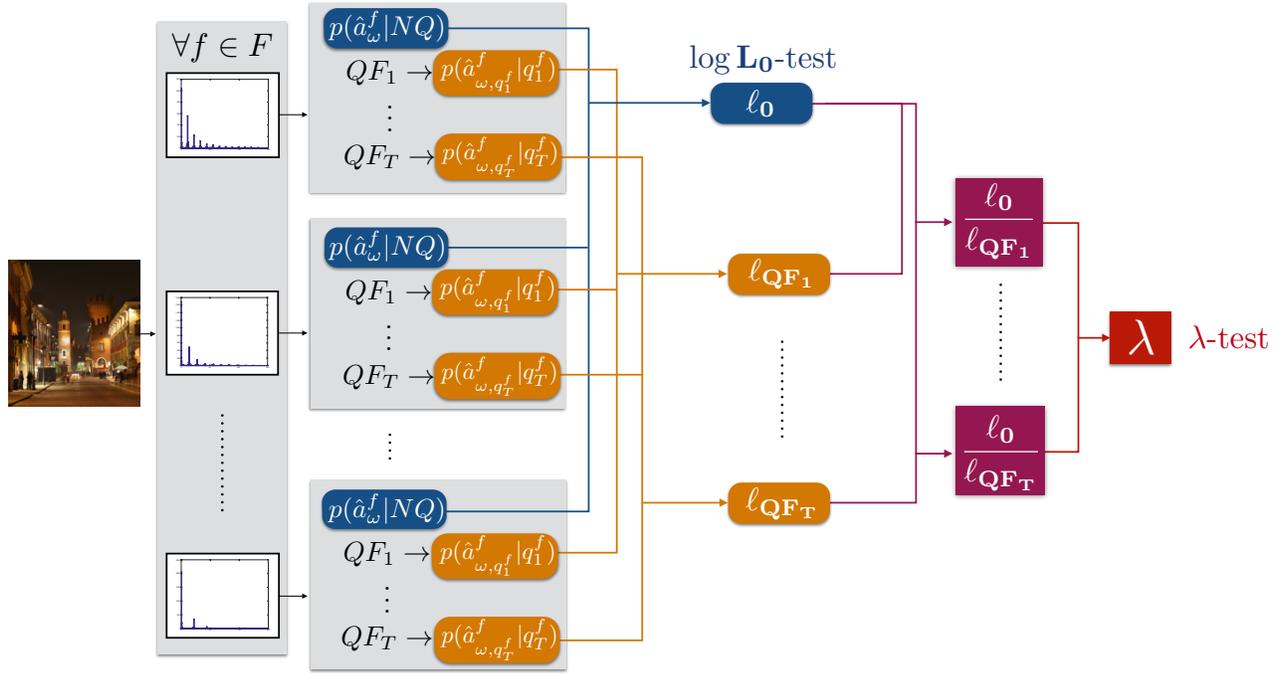


Fig. 4. Scheme of the proposed JPEG compression detectors. For each DCT frequency, the probabilities under the null and alternative hypotheses are computed (by means of the statistical derivations presented in Section IV) and used to obtain the related likelihood function values. The  $\log \mathbf{L}_0$ -test employs only the value of the likelihood function of the null hypothesis, while the  $\lambda$ -test gathers all the likelihood function values among the different alternative hypotheses.

where  $\mathcal{N}_2(\mu_{n,q}, \Sigma_{n,q})$  is the pdf of a real bivariate Gaussian r.v. of expected value  $\mu_{n,q}$  and covariance matrix  $\Sigma_{n,q}$ , and in our case

$$\mu_{n,q} = \begin{bmatrix} \Re(a_{n,q}) \\ \Im(a_{n,q}) \end{bmatrix},$$

$$\Sigma_{n,q} = \begin{bmatrix} \sigma_{W_r}^2 & Cov(W_r, W_i) \\ Cov(W_i, W_r) & \sigma_{W_i}^2 \end{bmatrix}.$$

It is worth observing that these findings are in accordance with the results we obtained in our previous work [21], where multiple JPEG compression is addressed. In fact, in that case we ultimately observed that BF coefficients of images compressed more than once are normally distributed in the complex plane. This opens the way to a generalized closed-form analysis in case of repeated quantization, that will be subject of future work.

## V. JPEG COMPRESSION DETECTION ALGORITHM

Given the statistical characterization of BF coefficients under both hypotheses of no previous quantization and quantization with a generic step, we can now exploit such results for JPEG compression detection.

In order to do this, we consider a pool of possible quality factors  $\{QF_1, \dots, QF_T\}$  and the corresponding quantization tables<sup>3</sup>, and define a hypothesis testing scheme where the null and alternative hypotheses, respectively, are given by

<sup>3</sup>In this work, we use quality factors from 1 to 100, for the sake of brevity, to indicate the related quantization tables, supposing to know their exact correspondence. However, the method can be generalized to any generic set of known quantization tables.

$\mathbf{H}_0$ : the image has never been compressed

$\mathbf{H}_1$ : the image has been compressed with a quality factor among  $QF_1, \dots, QF_T$ .

Please, note that we can consider the alternative hypothesis as composite, since it includes a set of possible parameters, the potential quality factors. Now, we propose two discriminative tests based on two different statistics, that will be explored in detail in Sections V-A and V-B: the  $\lambda$ -test, a hypothesis testing scheme encompassing statistical models of both null and alternative hypotheses, and the  $\log \mathbf{L}_0$ -test, based on a full statistical characterization of the Likelihood Function for the null hypothesis. Both of them automatically determine an acceptance region for the null hypothesis  $\mathbf{H}_0$  without the need of any training phase, thus providing a final binary output on the subject image (i.e., never compressed or compressed). The workflow of both algorithms is represented in Fig. 4 and is described in detail in the following.

### A. The $\lambda$ -test

The first test consists in the application of the Generalized Likelihood Ratio Test (GLRT) to the given problem. Let  $F$  be the set of DCT frequencies considered, then we can extract the DCT coefficients and compute for each  $f \in F$  the respective BF coefficient estimates. We will indicate such coefficients as  $\hat{a}_n^f$ , thus adding as superscript the reference to the DCT frequency. According to the procedure described in Section IV-A, we can straightforwardly obtain their respective probabilities  $p(\hat{a}_n^f | NQ)$  and (assuming statistical independence between DCT frequencies [20]) we can compute the likelihood function

value for the null hypothesis as

$$\ell_0 = \prod_{f \in F} p(\hat{a}_n^f | NQ). \quad (14)$$

Moreover, we need to provide a likelihood value for each parameter of the composite alternative hypothesis, hence for each quality factor considered. To this end, for each frequency  $f$  and quality factor  $Q_{F_i}$ , the corresponding quantization step  $q_{i,f}$  at that DCT frequency is retrieved. Similarly as the null hypothesis, we can compute the BF coefficient estimate  $\hat{a}_{n,q_{i,f}}$  and its probability value  $p(\hat{a}_{n,q_{i,f}} | q_{i,f})$  can be obtained as in Section IV-B, thus obtaining a likelihood function value for each  $Q_{F_i}$ :

$$\ell_{\mathbf{Q}_{F_i}} = \prod_{f \in F} p(\hat{a}_{n,q_{i,f}} | q_{i,f}). \quad (15)$$

Finally, according to the GLRT design, the discriminative statistic is given by

$$\lambda = \frac{\ell_0}{\max_{i \in \{1, \dots, T\}} \ell_{\mathbf{Q}_{F_i}}}. \quad (16)$$

Then, in order to provide a general method that does not rely on a training phase, we propose to obtain the final decision on the image as follows:

$$\begin{aligned} \mathbf{H}_0 \text{ is accepted} & \quad \text{if } \lambda \geq 1 \\ \mathbf{H}_0 \text{ is rejected} & \quad \text{if } \lambda < 1. \end{aligned}$$

Indeed, setting the threshold to 1 means that we reject the null hypothesis as soon as we find an alternative one which achieves a higher value of the likelihood function. Clearly, such choice is generally suboptimal, as a distribution model of  $\lambda$  would be required in order to fix an optimal threshold at a certain false alarm probability upper bound. However, we will see in Section VI that the results achieved in the experimental settings considered are comparable to the ones obtained with an optimal threshold, numerically derived from the data by means of a false alarm upper bound criterion.

### B. The $\log \mathbf{L}_0$ -test

As depicted in Fig. 4, the second proposed test relies on the value of the likelihood function for only the null hypothesis, which is thresholded in order to classify an image as never compressed or compressed. The test only requires the computation of  $\ell_0$ , thus resulting in a lower computational complexity with respect to the  $\lambda$ -test. Although less comprehensive, using only the statistical model in case of no compression allows for a full statistical characterization. Indeed, we can consider the likelihood function itself as a r.v.  $\mathbf{L}_0$  depending on the r.v.'s  $\hat{A}_n^f$  and study its distribution. Thus, we can reformulate expression (14) as

$$\mathbf{L}_0 = \prod_{f \in F} p(\hat{A}_n^f | NQ) \quad (17)$$

$$= \prod_{f \in F} 2M |\hat{A}_n^f| \exp(-M |\hat{A}_n^f|^2) \quad (18)$$

where  $M$  is the number of DCT blocks in the image.

Equivalently, we can consider its natural logarithm:

$$\log \mathbf{L}_0 = \#F \cdot \log(2M) + \sum_{f \in F} \log(|\hat{A}_n^f|) - M \sum_{f \in F} |\hat{A}_n^f|^2 \quad (19)$$

where  $\#F$  is the cardinality of  $F$ .

Then, by exploiting the fact that each  $\hat{A}_n$  is Rayleigh distributed (see IV-A) in case of no compression, we can derive the mean and variance of the random variable  $\log \mathbf{L}_0$  under the null hypothesis. We refer the reader to the Appendix B for the complete derivation and we report here the expression of the mean and variance of  $\log \mathbf{L}_0$

$$E\{\log \mathbf{L}_0\} = \#F \cdot \log(2M) - \#F \cdot \left( \frac{\log M}{2} + \frac{\gamma}{2} + 1 \right),$$

$$Var\{\log \mathbf{L}_0\} = \#F \cdot \frac{\pi^2}{24},$$

where  $\gamma$  is the Euler-Mascheroni constant [22].

We can notice that the statistical distribution of  $\log \mathbf{L}_0$  is properly scaled according to  $M$ , i.e., the number of DCT coefficients involved in the estimation of the Benford-Fourier coefficients, thus allowing for robustness when analyzing images of different size. Thanks to these results it is possible to design a threshold-based test on the value of the likelihood function by exploiting Chebyshev's inequality:

$$p\left(|\log \mathbf{L}_0 - E\{\log \mathbf{L}_0\}| \geq k \sqrt{Var\{\log \mathbf{L}_0\}}\right) \leq \frac{1}{k^2}, k \in \mathbb{Z}.$$

Thus, we can fix a significance level  $\alpha$  and by setting  $k = \pm \sqrt{1/\alpha}$  we have that the probability that  $\log \mathbf{L}_0$  deviates from  $E\{\log \mathbf{L}_0\}$  more than  $k$  times the standard deviation of  $\log \mathbf{L}_0$  is lower than  $\alpha$ . By considering that  $\log \mathbf{L}_0$  expresses the likelihood of the null hypothesis, we can define the threshold as follows

$$\tau(\alpha) = E\{\log \mathbf{L}_0\} - \sqrt{\frac{1}{\alpha}} \cdot \sqrt{Var\{\log \mathbf{L}_0\}},$$

and design a threshold-based test as follows:

$$\begin{aligned} \mathbf{H}_0 \text{ is accepted} & \quad \text{if } \log \ell_0 \geq \tau(\alpha) \\ \mathbf{H}_0 \text{ is rejected} & \quad \text{if } \log \ell_0 < \tau(\alpha). \end{aligned}$$

It is worth pointing out that  $\tau(\alpha)$  can be theoretically determined, thus avoiding again the need of any preliminary training on data.

## VI. EXPERIMENTAL RESULTS

We performed extensive tests in order to assess the effectiveness of the proposed approach. In the following, we describe the considered experimental settings and the results obtained on images coming from diverse sources, comparing the performance with existing forensic methods.

We consider images belonging to 4 different datasets and apply the methods to the full-size pictures. In particular, we used a subset of the database in [23] (**LIU**, 1000 images, 256x256), the UCID [24] database (**UCID**, 1338 images, 384x512), the Dresden database [25] (**DRESDEN**, 1488 images, different size ranging from  $3072 \times 2304$  to  $4352 \times 3264$ ) and the RAISE dataset [26] in its 2k version (**RAISE2K**, 2000 images,

different sizes ranging from  $3008 \times 2000$  to  $4928 \times 3264$ ). The different datasets and their size proportions are visually represented in Fig. 5.

Other seven state-of-the-art methods have been considered for comparisons, that can be grouped into two categories: the *threshold-based* techniques proposed in [7], [8], [10], [11] and [12] (denoted in the following as **BF SINGLE**, **BLOCK**, **DCT**, **NOISE** and **FACTOR**, respectively) and the *SVM-based* techniques proposed in [6] and [13] (denoted in

the following as **BF FFT** and **FSD**, respectively). They have been briefly described in Section II and all of them (except for **BF SINGLE** whose threshold is determined as in [7]) require some preliminary training phase, whose goal is to determine the discriminative threshold to be adopted for the threshold-based group and the SVM model for the SVM-based group. Therefore, each dataset was randomly divided into two parts, one for training and another one for testing. Thus, for each of the four datasets we have a training set TR and a training set

TABLE II

RESULTS OBTAINED WITH THE AUTOMATIC THRESHOLDS FOR THE  $\lambda$ -TEST, THE  $\log L_0$ -TEST AND THE **BF SINGLE** METHOD ON THE DIFFERENT DATASETS. PERFORMANCE LOSS WITH RESPECT TO THE USE OF  $P$ -OPTIMAL THRESHOLDS IS REPORTED.

LIU					UCID				
		$\tau_{\text{aut}}$	$\tilde{\tau}_{\text{TS}}$	Loss		$\tau_{\text{aut}}$	$\tilde{\tau}_{\text{TS}}$	Loss	
$\lambda$ -test	TN	97.0	99.3	2.3	$\lambda$ -test	TN	98.9	99.7	0.8
	TP	99.3	96.4	-2.7		TP	99.7	99.7	0.0
	ACC	96.7	96.9	0.1		ACC	99.5	99.7	0.1
	AUC	0.9957				AUC	0.9990		
$\log L_0$ -test	TN	100	99.5	-0.5	$\log L_0$ -test	TN	99.9	99.7	-0.2
	TP	94.3	97.6	3.3		TP	99.6	99.9	0.3
	ACC	95.3	97.9	2.6		ACC	99.6	99.8	0.2
	AUC	0.9918				AUC	1.0000		
<b>BF SINGLE</b>	TN	99.3	99.1	-0.2	<b>BF SINGLE</b>	TN	98.9	99.5	0.6
	TP	90.5	92.0	1.5		TP	99.2	99.3	0.1
	ACC	92.0	93.2	1.2		ACC	99.2	99.3	0.1
	AUC	0.9826				AUC	0.9997		

DRESDEN					RAISE2K				
		$\tau_{\text{aut}}$	$\tilde{\tau}_{\text{TS}}$	Loss		$\tau_{\text{aut}}$	$\tilde{\tau}_{\text{TS}}$	Loss	
$\lambda$ -test	TN	99.9	100	0.1	$\lambda$ -test	TN	100	99.9	-0.1
	TP	100	100	0.0		TP	99.8	99.9	0.1
	ACC	100	100	0.0		ACC	99.9	99.9	-0.0
	AUC	1.0000				AUC	1.0000		
$\log L_0$ -test	TN	99.3	100	0.7	$\log L_0$ -test	TN	95.4	100	4.6
	TP	100	100	0.0		TP	100	100	0.0
	ACC	99.9	100	0.1		ACC	99.2	100	0.8
	AUC	1.0000				AUC	1.0000		
<b>BF SINGLE</b>	TN	98.5	99.8	1.3	<b>BF SINGLE</b>	TN	98.0	100	2.0
	TP	99.9	99.8	-0.1		TP	100	100	0.0
	ACC	99.7	99.8	0.1		ACC	99.6	100	0.4
	AUC	0.9998				AUC	1.0000		



Fig. 5. Samples of test images.

TS. All the images have been compressed at quality factors  $\{90, 80, 70, 60, 50\}$ , features were extracted from the ones in the training set and used for determining the optimal threshold or for obtaining the SVM models.

We first evaluate our detectors designed in Section V and compare them with the group of threshold based methods in Section VI-A; then, in Section VI-B the SVM methods are considered for comparison. The analysis of locally forged areas is addressed in Section VI-C and, finally, Section VI-D evaluates the ability of the  $\log \mathbf{L}_0$ -test detector in rejecting the null hypothesis in case of double compressed and then decompressed images.

#### A. Comparison with the threshold-based methods

In this section we test our proposed detectors by using the automatic thresholds introduced in Section V and compare them to the five threshold-based methods.

First, we illustrate our definition of optimality of a threshold, according to the Neyman-Pearson criterion. Given a set of images  $S$  and a certain threshold-based test, the discriminative value can be varied to obtain a Receiver Operating Curve (ROC). The  $P$ -optimal threshold  $\tilde{\tau}_S(P)$  on  $S$  is defined as

follows: among the ones leading to a false positive rate on  $S$  not higher than  $P$ , we identify the set of discriminative values that yield the highest true positive rate; then, we find in this set the value  $\tilde{\tau}_S(P)$  that yields the lowest false alarm rate. In this section, we always consider  $P$  equal to 1% and, for the sake of brevity, we omit the dependence of  $\tilde{\tau}_S$  on  $P$ . Other values of  $P$  will be considered in Section VI-B and will be explicitly reported.

As a first test, we consider our two proposed detectors, the  $\lambda$ -test and the  $\log \mathbf{L}_0$ -test, together with and the **BF SINGLE** method<sup>4</sup>, as all of them allow automatic thresholds. As no training is needed in this case, we consider only TS for each dataset. Then, the images belonging to TS of each dataset are tested in six different versions: negative samples are uncompressed images and positive samples are the same images compressed with quality factor 50, 60, 70, 80 and 90. The three methods are applied in two different ways: first the automatic thresholds are computed (where the significance for the  $\log \mathbf{L}_0$ -test and **BF SINGLE** is fixed to 0.01) and denoted with  $\tau_{\text{aut}}$ ; then, the  $P$ -optimal threshold  $\tilde{\tau}_{\text{TS}}$  on TS is employed.

In Table II, we report the results obtained by applying a 10-fold cross validation strategy (i.e., at each validation step 1/10 of the images in each dataset is used for testing and the rest for training), which has been adopted through the whole subsection. Different thresholds used are arranged column-wise, while the three tests are reported row-wise. We are interested in observing the results of the detectors but also in assessing the performance loss when using  $\tau_{\text{aut}}$  instead of  $\tilde{\tau}_{\text{TS}}$ . In other words, we can measure how much we “lose” by employing the automatic thresholds instead of the  $P$ -optimal ones. Then, for each case results are reported in terms of true negative rate **TN**, true positive rate **TP**, accuracy **ACC** and Area Under the Curve **AUC**. Finally, the last column contains

<sup>4</sup>For all the tests, we fixed the value of  $n = 4$ , coherently with our first approach in [7]. Moreover, for the  $\lambda$ -test and  $\log \mathbf{L}_0$ -test we limited the analysis to the first 10 DCT frequencies in zig-zag order, in order to avoid a high number of null coefficients.

TABLE III

RESULTS OF STATE-OF-THE-ART THRESHOLD-BASED METHODS IN THE HOMOGENEOUS SCENARIO. PERFORMANCE LOSS IN USING THE  $P$ -OPTIMAL THRESHOLDS OF THE TRAINING SET INSTEAD OF THE TESTING SET FOR EACH DATASET IS REPORTED.

		LIU			UCID			DRESDEN			RAISE2K		
		$\tilde{\tau}_{\text{TR}}$	$\tilde{\tau}_{\text{TS}}$	Loss									
<b>BLOCK</b>	<b>TN</b>	99.1	99.0	-0.1	99.5	99.4	-0.1	100	100	0.0	100	99.9	-0.1
	<b>TP</b>	78.1	80.7	2.6	99.0	99.2	0.2	99.9	100	0.1	99.9	100	0.1
	<b>ACC</b>	81.6	83.8	2.2	99.0	99.2	0.2	99.9	100	0.1	99.9	99.9	0.0
	<b>AUC</b>	0.9695			0.9996			1.0000			1.0000		
	<b>AUC</b>	0.9695			0.9996			1.0000			1.0000		
<b>DCT</b>	<b>TN</b>	98.8	99.3	0.5	99.1	99.4	0.3	100	100	0.0	99.1	99.5	0.4
	<b>TP</b>	97.7	98.0	0.3	98.0	98.2	0.2	100	100	0.0	99.9	99.7	-0.2
	<b>ACC</b>	97.9	98.3	0.4	98.2	98.4	0.2	100	100	0.0	99.8	99.7	0.1
	<b>AUC</b>	0.9988			0.9966			1.0000			0.9998		
	<b>AUC</b>	0.9988			0.9966			1.0000			0.9998		
<b>NOISE</b>	<b>TN</b>	98.9	99.2	0.3	98.7	99.2	0.5	99.9	99.9	0.0	99.2	99.4	0.2
	<b>TP</b>	97.8	98.2	0.4	97.5	97.0	-0.5	99.8	100	0.2	78.8	73.2	-5.6
	<b>ACC</b>	97.9	98.3	0.4	97.7	97.4	-0.3	99.8	99.9	0.1	82.2	77.6	-4.6
	<b>AUC</b>	0.9960			0.9967			1.0000			0.9925		
	<b>AUC</b>	0.9960			0.9967			1.0000			0.9925		
<b>FACTOR</b>	<b>TN</b>	98.9	99.7	0.8	99.3	99.6	0.3	100	100	0.0	100	100	0.0
	<b>TP</b>	99.8	99.8	0.0	99.7	99.7	0.0	99.9	99.9	0.0	99.9	100	0.1
	<b>ACC</b>	99.7	99.8	0.1	99.6	99.7	0.1	100	100	0.0	99.9	100	0.1
	<b>AUC</b>	0.9991			0.9997			1.0000			1.0000		
	<b>AUC</b>	0.9991			0.9997			1.0000			1.0000		

the performance loss between the two thresholds. Thus, if  $\text{TN}(\tilde{\tau}_{\text{TS}})$  and  $\text{TN}(\tau_{\text{aut}})$  are the true negative rates obtained with  $\tilde{\tau}_{\text{TS}}$  and  $\tau_{\text{aut}}$ , respectively, the loss is computed as

$$\text{TN}(\tilde{\tau}_{\text{TS}}) - \text{TN}(\tau_{\text{aut}}),$$

where a positive (negative) loss means a performance decrease (increase). The same holds for **TP** and **ACC**.

It is worth noticing that the true negative rate **TN** is the complementary of the false positive rate (i.e., the minimum value for optimal thresholds is at least 99.0%) and it is used equivalently for uniformly defining the performance loss; moreover, the positive samples are more numerous than negative samples, thus, the accuracy value is more influenced by the **TP** rather than the **TN** value. In Table II, we also plot below each dataset the ROC curve of the two proposed detectors for one TR – TS partition, in order to highlight the displacement between results obtained with  $\tilde{\tau}_{\text{TS}}$  (orange circle) and  $\tau_{\text{aut}}$  (yellow circle).

We can notice that both detectors yield good accuracies

values in any case and outperform **BF SINGLE**, which is based on a single Benford-Fourier coefficient. Moreover, the performance loss is generally limited, in the sense that either the **TP** and the **TN** rates are not significantly compromised. In particular, it is interesting to notice that the automatic threshold for the  $\log L_0$ -test, determined with the false alarm probability upper bound fixed to 0.01, indeed yields a false positive rate below 1% in every case, with the exception of the RAISE2K dataset where the false positive rate raise to 4.3%. By selecting the outlier images from this database and exploring their statistics, we have noticed that DCT coefficients in uncompressed images (that are supposed to be GG distributed) usually present some anomalies, which cause deviations with respect to the models predicted in Section IV-A and disappear when the image is decimated by a factor 2 in each dimension. We conjecture such phenomenon is due to specific capture settings (all outlier images have been taken with the same camera) coupled with the high resolution, which potentially increases the correlation among image blocks.

TABLE IV

PERFORMANCE OF STATE-OF-THE-ART THRESHOLD-BASED METHODS IN THE NON-HOMOGENEOUS SCENARIO. PERFORMANCE LOSS IN USING THE  $P$ -OPTIMAL THRESHOLDS OF THE TRAINING SET INSTEAD OF THE TESTING SET FOR EACH DATASET PAIR IS REPORTED. EACH CELL CONTAINS THE LOSS IN TERMS OF TRUE NEGATIVE AND TRUE POSITIVE RATE, RESPECTIVELY. THE ROC PLOTS AND THE HISTOGRAMS OF POSITIVE SAMPLE FEATURES FOR THE TRAINING AND TESTING SET REFER TO THE CASE REPORTED IN BOLD IN THE CORRESPONDING TABLE.

		BLOCK				
		TR	LIU	UCID	DRESDEN	RAISE2K
TS	LIU	-0.1/2.6	65.8/-19.2	95.9/-19.3	99.0/-19.3	<b>99.3/-0.8</b>
	UCID	-0.6/12.0	-0.1/0.2	77.5/-0.8	<b>99.3/-0.8</b>	1.4/0.0
	DRESDEN	0.0/1.7	0.0/0.4	0.0/0.1	0.0/0.0	0.1/0.0
	RAISE2K	0.0/10.7	0.0/1.1	0.0/0.0	0.0/0.0	0.1/0.0

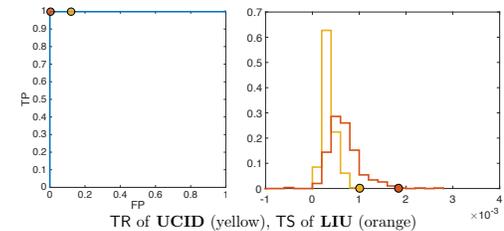
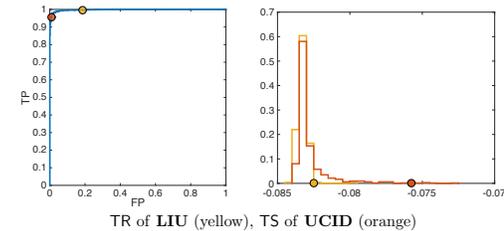
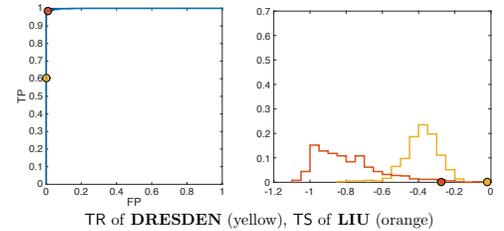
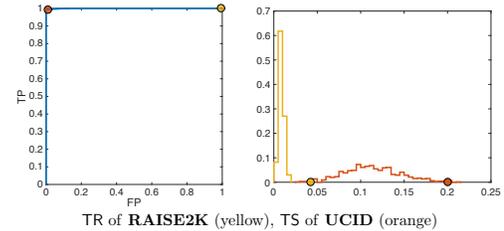
		DCT				
		TR	LIU	UCID	DRESDEN	RAISE2K
TS	LIU	0.5/0.3	-0.6/5.2	<b>-0.7/26.8</b>	-0.6/8.3	-0.4/1.0
	UCID	6.9/-0.9	0.2/0.2	-0.6/12.0	-0.4/1.0	0.1/0.0
	DRESDEN	10.3/0.0	0.1/0.0	0.0/0.0	0.0/0.0	0.1/0.0
	RAISE2K	21.3/-0.3	1.4/-0.3	-0.3/3.7	0.5/-0.2	0.5/-0.2

		NOISE				
		TR	LIU	UCID	DRESDEN	RAISE2K
TS	LIU	0.3/0.4	-0.8/13.3	-0.8/25.7	-0.8/54.1	-0.8/32.4
	UCID	<b>13.7/-2.6</b>	0.5/-0.5	-0.3/4.0	-0.8/32.4	-0.1/5.7
	DRESDEN	42.3/-0.0	0.1/-0.0	0.0/0.1	0.0/0.1	0.2/5.5
	RAISE2K	49.2/-26.8	8.9/-26.4	3.6/-23.9	0.2/5.5	0.2/5.5

		FACTOR				
		TR	LIU	UCID	DRESDEN	RAISE2K
TS	LIU	0.8/0.0	<b>25.6/-0.0</b>	96.5/-0.1	91.2/-0.1	87.3/-0.2
	UCID	-0.4/0.4	0.3/0.1	80.4/-0.2	87.3/-0.2	0.0/0.0
	DRESDEN	0.0/0.5	0.0/0.1	0.0/0.0	0.0/0.0	0.0/0.0
	RAISE2K	0.0/0.1	0.0/0.0	0.0/0.0	0.0/0.0	0.0/0.0



Further investigation on these anomalies will be subject of future work.

With the same spirit, we perform a similar test for the 4 state-of-the-art threshold-based methods. Since no explicit models for the statistics they use are available, in order to find a reference threshold, we used the training sets TR of the datasets. Thus, for every method and dataset we first determine numerically by exhaustive search the threshold  $\tilde{\tau}_{TR}$  that is  $P$ -optimal on TR, and then we apply such threshold on TS. We report the loss performance loss with respect to the specific  $\tilde{\tau}_{TS}$ , so that results refer to the very same set of images of Table II.

We first consider an experimental setting where both TR and TS come from the same dataset. For instance, threshold  $\tilde{\tau}_{TR}$  obtained from TR of **UCID** is tested on TS of **UCID**. For the second experimental setting, the image set TR from a different dataset is used to obtain the reference threshold  $\tilde{\tau}_{TR}$  which is tested on TS from a different dataset. We will indicate these settings as the *homogeneous* and *non-homogeneous* scenarios, respectively.

Results for the homogeneous scenario are reported in Table III. In this case the threshold-based methods provide results that are comparable and in some cases superior to our detectors.

However, when moving to the non-homogeneous scenario we observe that the performance of the threshold-based detectors is no longer reliable. In Table IV, we report the results for all the training/testing dataset combinations for each threshold-based method. Different training sets are arranged column-wise, while testing sets are placed row-wise. Due to space constraints, each cell contains only two values, corresponding to the loss in terms of **TN** and **TP**, respectively. Although in some cases thresholds are relatively robust, we can notice that the performance often drops in one of these two indicators.

In order to explore the causes of such underperformance, in one selected case for each dataset (corresponding to bold numbers in the table) we also plot the ROC curve on the testing

set for one TR – TS partition. The performance obtained with  $\tilde{\tau}_{TS}$  and  $\tilde{\tau}_{TR}$  is marked with orange and yellow circles, respectively. Moreover, for the same cases we report the normalized histogram of the statistic values of the specific method<sup>5</sup> for negative samples (uncompressed images) from TS (orange line) and TR (yellow line). The corresponding optimal thresholds  $\tilde{\tau}_{TS}$  (orange circle) and  $\tilde{\tau}_{TR}$  (yellow circle) are also plotted. We can observe that the two histograms in the selected cases are substantially different and hardly predictable, thus causing the optimal thresholds to be far apart from each other. Then, although the statistics used are actually discriminative ( $\tilde{\tau}_{TS}$  yields very good results on TS), the choice of the threshold when dealing with diverse data still represents a significant issue for threshold-based approaches. On the other hand, the  $\lambda$ -**test** and  $\log \mathbf{L}_0$ -**test** come with an automatic thresholds according to the images tested and, thanks to the robust and adaptive models, they achieve the stable results across the different datasets reported in Table II. In other words, as long as the data follow the GG assumption, as natural images do, there is no dataset-dependency.

### B. Comparison with SVM-based methods

We now compare our method with the SVM-based techniques. In this phase, the SVM models are obtained from the TR as suggested in the original papers. With respect to the scenarios introduced in Section VI-A, in these experiments we adopt the homogenous setting, i.e., a classifier for each dataset is obtained and tested on the remaining images of the same dataset.

To perform a fair comparison, we consider the true negative rate obtained by the considered SVM method on every TR (denoted in each case as  $\mathbf{TN}_{SVM}^{TR}$ ) and determine the thresholds for the  $\lambda$  and  $\mathbf{L}_0$  statistics as the  $(1 - \mathbf{TN}_{SVM}^{TR})$ -optimal

<sup>5</sup>For the sake of consistency, in the **DCT** and **NOISE** methods we consider the opposite of the statistic values.

TABLE V  
COMPARISON OF THE  $\lambda$ -**test** AND THE  $\log \mathbf{L}_0$ -**test** TO [6] WITH TRUE NEGATIVE RATE MATCHING.

		LIU					UCID						
		90	80	70	60	50	90	80	70	60	50		
<b>BF FFT</b> , $\mathbf{TN}_{SVM}^{TR} = 74.8$	<b>TN</b>			74.2			<b>BF FFT</b> , $\mathbf{TN}_{SVM}^{TR} = 92.1$	<b>TN</b>			93.6		
	<b>TP</b>	99.4	99.8	96.6	98.8	98.2		<b>TP</b>	99.5	99.1	98.0	99.0	98.2
	<b>ACC</b>	86.8	87.0	85.4	86.5	86.2		<b>ACC</b>	96.6	96.3	95.8	96.3	95.9
$\lambda$ - <b>test</b> , $\tilde{\tau}_{TR}(1 - \mathbf{TN}_{SVM}^{TR})$	<b>TN</b>			75.2			$\lambda$ - <b>test</b> , $\tilde{\tau}_{TR}(1 - \mathbf{TN}_{SVM}^{TR})$	<b>TN</b>			96.0		
	<b>TP</b>	97.2	98.8	99.4	100	99.2		<b>TP</b>	99.7	99.7	99.9	99.9	99.9
	<b>ACC</b>	86.2	87.0	87.3	87.6	87.2		<b>ACC</b>	97.8	97.8	97.9	97.9	97.9
$\log \mathbf{L}_0$ - <b>test</b> , $\tilde{\tau}_{TR}(1 - \mathbf{TN}_{SVM}^{TR})$	<b>TN</b>			78.2			$\log \mathbf{L}_0$ - <b>test</b> , $\tilde{\tau}_{TR}(1 - \mathbf{TN}_{SVM}^{TR})$	<b>TN</b>			92.4		
	<b>TP</b>	96.0	99.8	99.6	99.6	99.8		<b>TP</b>	99.9	99.9	99.9	100	100
	<b>ACC</b>	87.1	89.0	88.9	88.9	89.0		<b>ACC</b>	96.1	96.1	96.1	96.2	96.2
		DRESDEN					RAISE2K						
		90	80	70	60	50	90	80	70	60	50		
<b>BF FFT</b> , $\mathbf{TN}_{SVM}^{TR} = 94.9$	<b>TN</b>			96.0			<b>BF FFT</b> , $\mathbf{TN}_{SVM}^{TR} = 93.8$	<b>TN</b>			95.3		
	<b>TP</b>	100	99.7	97.3	99.9	99.8		<b>TP</b>	99.8	99.8	99.5	98.9	99.6
	<b>ACC</b>	98.0	97.9	96.6	97.6	97.9		<b>ACC</b>	97.6	97.6	97.4	97.1	97.4
$\lambda$ - <b>test</b> , $\tilde{\tau}_{TR}(1 - \mathbf{TN}_{SVM}^{TR})$	<b>TN</b>			100			$\lambda$ - <b>test</b> , $\tilde{\tau}_{TR}(1 - \mathbf{TN}_{SVM}^{TR})$	<b>TN</b>			96.1		
	<b>TP</b>	100	100	99.7	100	100		<b>TP</b>	99.9	100	100	100	100
	<b>ACC</b>	100	100	99.9	100	100		<b>ACC</b>	98.0	98.0	98.0	98.0	98.0
$\log \mathbf{L}_0$ - <b>test</b> , $\tilde{\tau}_{TR}(1 - \mathbf{TN}_{SVM}^{TR})$	<b>TN</b>			99.9			$\log \mathbf{L}_0$ - <b>test</b> , $\tilde{\tau}_{TR}(1 - \mathbf{TN}_{SVM}^{TR})$	<b>TN</b>			94.6		
	<b>TP</b>	100	100	100	100	100		<b>TP</b>	100	100	100	100	100
	<b>ACC</b>	99.9	99.9	99.9	99.9	99.9		<b>ACC</b>	97.3	97.3	97.3	97.3	97.3

TABLE VI  
COMPARISON OF THE  $\lambda$ -**test** AND THE  $\log \mathbf{L}_0$ -**test** TO [13] WITH TRUE NEGATIVE RATE MATCHING.

LIU						UCID							
		90	80	70	60	50		90	80	70	60	50	
<b>FSD</b> , $\text{TN}_{\text{SVM}}^{\text{TR}} = 92.8$	<b>TN</b>			92.6			<b>FSD</b> , $\text{TN}_{\text{SVM}}^{\text{TR}} = 97.6$	<b>TN</b>			97.5		
	<b>TP</b>	87.4	97.8	94.8	87.8	82.8		<b>TP</b>	96.4	99.1	98.6	96.4	95.5
	<b>ACC</b>	90.0	95.2	93.7	90.2	87.7		<b>ACC</b>	96.9	98.3	98.0	96.9	96.5
$\lambda$ - <b>test</b> , $\tilde{\tau}_{\text{TR}}(1 - \text{TN}_{\text{SVM}}^{\text{TR}})$	<b>TN</b>			92.6			$\lambda$ - <b>test</b> , $\tilde{\tau}_{\text{TR}}(1 - \text{TN}_{\text{SVM}}^{\text{TR}})$	<b>TN</b>			98.8		
	<b>TP</b>	87.6	98.6	98.6	99.4	99.0		<b>TP</b>	99.4	99.6	99.9	99.9	99.9
	<b>ACC</b>	90.1	95.6	95.6	96.0	95.8		<b>ACC</b>	99.1	99.2	99.3	99.9	99.3
$\log \mathbf{L}_0$ - <b>test</b> , $\tilde{\tau}_{\text{TR}}(1 - \text{TN}_{\text{SVM}}^{\text{TR}})$	<b>TN</b>			94.0			$\log \mathbf{L}_0$ - <b>test</b> , $\tilde{\tau}_{\text{TR}}(1 - \text{TN}_{\text{SVM}}^{\text{TR}})$	<b>TN</b>			99.5		
	<b>TP</b>	93.8	99.8	99.4	99.6	99.8		<b>TP</b>	99.9	99.9	99.9	99.7	100
	<b>ACC</b>	93.0	96.9	96.7	96.8	96.9		<b>ACC</b>	99.6	99.7	99.7	99.6	99.8
DRESDEN						RAISE2K							
		90	80	70	60	50		90	80	70	60	50	
<b>FSD</b> , $\text{TN}_{\text{SVM}}^{\text{TR}} = 99.7$	<b>TN</b>			97.5			<b>FSD</b> , $\text{TN}_{\text{SVM}}^{\text{TR}} = 93.8$	<b>TN</b>			99.2		
	<b>TP</b>	98.4	100	98.5	91.1	89.1		<b>TP</b>	86.3	96.8	98.5	94.4	91.4
	<b>ACC</b>	97.9	98.7	98.0	94.3	93.3		<b>ACC</b>	92.8	98.0	98.9	96.8	95.3
$\lambda$ - <b>test</b> , $\tilde{\tau}_{\text{TR}}(1 - \text{TN}_{\text{SVM}}^{\text{TR}})$	<b>TN</b>			100.0			$\lambda$ - <b>test</b> , $\tilde{\tau}_{\text{TR}}(1 - \text{TN}_{\text{SVM}}^{\text{TR}})$	<b>TN</b>			99.9		
	<b>TP</b>	100	100	99.7	100	100		<b>TP</b>	99.9	99.9	100	100	100
	<b>ACC</b>	100	100	99.9	100	100		<b>ACC</b>	99.8	99.9	99.9	99.9	99.9
$\log \mathbf{L}_0$ - <b>test</b> , $\tilde{\tau}_{\text{TR}}(1 - \text{TN}_{\text{SVM}}^{\text{TR}})$	<b>TN</b>			99.9			$\log \mathbf{L}_0$ - <b>test</b> , $\tilde{\tau}_{\text{TR}}(1 - \text{TN}_{\text{SVM}}^{\text{TR}})$	<b>TN</b>			100		
	<b>TP</b>	100	100	100	100	100		<b>TP</b>	99.9	99.8	100	100	100
	<b>ACC</b>	99.9	99.9	99.9	99.9	99.9		<b>ACC</b>	99.9	99.9	100	100	100

threshold on TR, according to the definition given in Section VI-A. In other words, we allow the false positive rate of our method to be at most equal to the one obtained by the SVM-based method on the same TR. We indicate such threshold as  $\tilde{\tau}_{\text{TR}}(1 - \text{TN}_{\text{SVM}}^{\text{TR}})$ , in accordance with the previous section. The SVM classifiers and thresholds obtained are then applied to the corresponding TS.

In Tables V and VI, the performance of the  $\lambda$ -**test** and  $\log \mathbf{L}_0$ -**test** versus **BF FFT** and **FSD**, respectively, are reported. To get a deeper understanding of the results, we consider the different quality factors separately and arrange them column-wise.

We can observe that both proposed detectors yield higher accuracies and at least one of them outperforms the considered SVM-based methods in every case. However, it is worth noticing that this experimental setting is adopted in order to fairly compare our proposed methods with SVM classifiers by matching the true negative rates in the training set, while the possibility of automatically determining the threshold for the two proposed tests, which is one of the strong advantages of our methods, is not exploited.

### C. Local analysis

The results presented so far were obtained on full frame images, while the proposed method can also be applied locally and reveal inconsistencies within the same picture, thus suggesting that a manipulation occurred. This happens, for instance, when a JPEG image is manipulated by adding one or more parts either coming from uncompressed images or processed so that compression traces disappear. If the image is finally saved in uncompressed format, the original part will present traces of compression while the added part will likely behave as uncompressed in terms of DCT statistics.

In order to synthetically reproduce such situation, we considered images from the Dresden dataset and compressed

them with a random quality factor from 50 to 100. Then, we replaced a square part of  $1536 \times 2048$  pixels with its original uncompressed version.

For space constraints and in light of the results obtained in Section VI-A, we use the  $\log \mathbf{L}_0$ -test, which proved to be accurate on small pictures and requires a lower computational complexity with respect to the  $\lambda$ -test. We performed a local analysis by applying the test to non-overlapping windows and by classifying each of them as compressed or non compressed according to a threshold on  $\log \mathbf{L}_0$ , which is computed as function of the number of DCT blocks in each window and the desired significance level (set to 0.01). We have repeated this operation for three different window sizes and the results, reported in Table VII, show that the proposed method achieves rather good accuracies also when the window size is  $128 \times 128$ , also thanks to its size-adaptive nature. However, when the size of the window further decreases the detection rate decreases as well, generally faster than methods in [11] and [12] which prove to be more suitable for a fine-grained block analysis. The reason for that is two-fold: fewer DCT coefficients are used in the estimation and, being drawn from smaller patches in high resolution images, likely are more correlated to each other, thus not totally fulfilling the independence assumption in the model.

TABLE VII  
CLASSIFICATION RESULTS OF THE LOCAL ANALYSIS THROUGH ALL THE IMAGES WITH THE  $\log \mathbf{L}_0$ -**test**.

	512 × 512	256 × 256	128 × 128
<b>TN</b>	99.5	99.8	100.0
<b>TP</b>	95.7	89.6	79.0
<b>ACC</b>	97.0	92.9	85.7

As further demonstration of the effectiveness of our approach also in non-synthetic settings, we report in Fig. 6 an example obtained on a visually compelling manipulated image. In this case, an object coming from a different picture has



Fig. 6. A forged image is obtained by composing two JPEG images: an object from a JPEG image (downloaded from the web) has been superimposed on another JPEG background image (captured with a Nexus 5 smartphone), and the final composite image has been saved in an uncompressed format. The inserted object underwent strong modifications due to visual requirements, thus essentially losing traces of potential previous compressions, while the background image still contains those traces. The local value of  $\log \mathbf{L}_0$  has then been computed. Overlapping windows of  $120 \times 120$  pixels with a stride equal to 2 pixels have been considered. The  $\log \mathbf{L}_0$  map shows that such local analysis correctly identifies the different statistical properties within the image.

TABLE VIII

RESULTS OF THE  $\log \mathbf{L}_0$ -TEST METHOD WHEN TWICE COMPRESSED IMAGES ARE CONSIDERED. PERFORMANCE LOSS WITH RESPECT TO THE USE OF  $P$ -OPTIMAL THRESHOLDS IS REPORTED.

		LIU			UCID			DRESDEN			RAISE2K		
		$\tau_{\text{aut}}$	$\hat{\tau}_{\text{TS}}$	Loss									
$\log \mathbf{L}_0$ -test	TN	100	99.4	-0.6	99.9	99.9	0.0	99.7	99.8	0.1	94.9	100	4.6
	TP	98.6	99.1	0.5	100	100	0.0	100	100	0.0	100	100	0.0
	ACC	98.8	99.2	0.4	100	100	0.0	99.9	99.9	0.0	99.1	99.9	0.8
	AUC	0.9965			1.0000			1.0000			1.0000		

been inserted into an existing JPEG image and the composite image has been saved in an uncompressed format. We can see from the map that the background is correctly characterized by lower values of  $\log \mathbf{L}_0$ .

#### D. Rejection of the null hypothesis in case of double compressed images

The ability of distinguishing once-compressed from never compressed images does not imply that the same detector would distinguish once-compressed from twice compressed images. However, it is possible that the  $\log \mathbf{L}_0$ -test, which does not rely on an explicit model for the alternative hypothesis, is able to reject the null hypothesis also when a twice-compressed (and decompressed) image is tested.

In fact, we have carried out experiments in the same setting as Section VI-A but the positive images were further compressed with quality factor 70. We then applied the  $\log \mathbf{L}_0$ -test as we did in Section VI-A, where positive images are now compressed twice. Results are reported in Table VIII and show very high true positive rates (also for images compressed twice with the same quality factor 70), thus confirming the ability of the  $\log \mathbf{L}_0$ -test of correctly rejecting the null hypothesis.

## VII. CONCLUSIONS AND DISCUSSION

We have proposed a novel statistical analysis of BF coefficients, which is exploited to design two threshold-based forensic detectors able to reveal the traces of JPEG compression in digital images stored in uncompressed formats. Thanks to the statistical models developed, for both detectors it is possible to automatically determine the threshold to be used, thus avoiding any preliminary training phase on data. The experiments performed on real images of different sizes and

source cameras show that such an approach yields to results that are comparable to existing methods and has the additional advantage that there is no need for further training/calibration of the threshold when testing different kind of images. Indeed, the proposed detectors are not affected by dataset-dependency issues, as the automatically determined thresholds lead to robust and stable results among different datasets. Moreover, the  $\log \mathbf{L}_0$ -test does not require the knowledge of the possible quantization tables used, but simply rejects the hypothesis of no compression when a deviation from the expected statistics is observed.

However, a number of aspects represent future research directions, like the extension of our methods to other compression standards, as well as their evaluation in a high quality scenario or adversarial perspective.

Indeed, we assessed that the proposed detectors lose effectiveness when the compression is performed at very high quality. This can be explained by the fact that high factors ( $\geq 94/95$ ) involve very small quantization steps (often equal to 1 for the DCT frequencies used), which result in very similar statistical properties between never compressed and compressed/dequantized images so that the model assumed in Section IV-B might not be accurate. Given the good performance obtained by techniques like [11] also in this scenario, a proper method combination and a smarter selection of the DCT frequencies could represent a viable option for future developments, as already explored in [27] in the single vs double compression scenario. Moreover, while counterforensics methods restoring the Generalized Gaussian distribution [28] [15] would likely compromise the detection accuracy, additional compound statistical properties could be studied to detect and counter JPEG anti-forensics, in a similar way as in [29] and [30].

Finally, the encouraging results obtained in Section VI-D, as well as in the work [21], confirm the effectiveness of BF coefficients in characterizing DCT coefficients. This suggests that future work should be devoted to the analysis of more complex scenarios involving multiple JPEG compression and other processing operations [31], [32].

#### ACKNOWLEDGEMENT

We would like to thank Dr. Bin Li and Dr. Janquiang Yang for kindly providing us with the code of their methods in [11] and [12], respectively.

#### APPENDIX A DERIVATION OF $\sigma_{W_{r,i}}^2$

In order to obtain the variance of  $W_{r,i}$ , we need to consider the pdf of  $Z_q$  and study the real and imaginary parts of the r.v.  $e^{-j2\pi n \log_{10} Z_q}$ , i.e., the r.v.'s  $C \doteq \cos(2\pi n \log_{10} Z_q)$  and  $S \doteq -\sin(2\pi n \log_{10} Z_q)$ , respectively. For the sake of simplicity, in the following analysis we drop the subscript  $q$  of  $Z$ . For deriving the pdf of  $C$ , the following r.v. transformations need to be applied

$$Z \xrightarrow{\log_{10}} Z' \xrightarrow{\cdot 2\pi n} Z'' \xrightarrow{\cos} C$$

and the same happens for  $S$  by applying  $-\sin(\cdot)$  as last transformation.

Since they are monotonic, the first two transformations can be treated with the formula

$$f_Y(y) = f_X(h^{-1}(y)) \cdot \left| \frac{\partial}{\partial y} h^{-1}(y) \right|, \quad Y = h(X), X \sim f_X(x)$$

and we obtain

$$f_{Z''}(z'') = \mathcal{L}(10^{\frac{z''}{2\pi n}}) \cdot 10^{\frac{z''}{2\pi n}} \cdot \frac{\log 10}{2\pi n}.$$

The third transformation is not monotonic and a generalization of formula (A) should be used [33]:

$$f_Y(y) = \sum_{\{x|h(x)=y\}} \frac{f_X(x)}{\left| \frac{\partial}{\partial x} h(x) \right|} \quad (20)$$

So, if we define  $I_{Z''} = [2\pi n \log_{10}(q - q/2), 2\pi n \log_{10}(q + q/2)]$ , we have that

$$\begin{aligned} f_C(c) &= \sum_{\{z''|\cos(z'')=c\} \cap I_{Z''}} \frac{f_{Z''}(z'')}{|-\sin z''|} \\ &= \sum_{D_c} \frac{e^{-\frac{\sqrt{2}}{\sigma} |10^{\frac{z''}{2\pi n}} - q|} 10^{\frac{z''}{2\pi n}} \log 10}{\sigma \sqrt{2} N_{\sigma,q} 2\pi n} \frac{1}{\sqrt{1-c^2}} \end{aligned}$$

where

$$\begin{aligned} D_c &= \{z''|\cos(z'')=c\} \cap I_{Z''} \\ &= \cup_{i \in \mathbb{Z}} \{2\pi i + t, 2\pi(i+1) - t\} \cap I_{Z''}, \\ t &\doteq \arccos(c). \end{aligned}$$

By definition, the variance of  $C$  is given by

$$\begin{aligned} \sigma_C^2 &= \int_{-1}^1 c^2 f_C(c) dc - \mu_C^2 \\ &= \int_{-1}^1 c^2 \sum_{z'' \in D_c} \frac{e^{-\frac{\sqrt{2}}{\sigma} |10^{\frac{z''}{2\pi n}} - q|} 10^{\frac{z''}{2\pi n}} \log 10}{\sigma \sqrt{2} N_{\sigma,q} 2\pi n} \frac{\log 10}{\sqrt{1-c^2}} dc - \Re(a_{n,q})^2. \end{aligned} \quad (21)$$

The variance  $\sigma_S^2$  of  $S$  is obtained similarly by replacing  $c$  with  $s$  in the first integral and recalling that  $\mu_S = \Im(a_{n,q})$ . Finally, in order to get the variance of the real and imaginary parts of  $W_{0,q} = \hat{A}_{n,q} - a_{n,q}$  we should consider that it is a shift (i.e., has the same variance) of  $\hat{A}_{n,q}$ , the sample mean of  $e^{-j2\pi n \log_{10} Z_q}$  computed from  $M_q$  samples. By applying the Central Limit Theorem on real and imaginary parts, we have:

$$\sigma_{W_r}^2 = \frac{\sigma_C^2}{M_q}, \quad \sigma_{W_i}^2 = \frac{\sigma_S^2}{M_q}. \quad (23)$$

#### APPENDIX B MEAN AND VARIANCE OF $\log \mathbf{L}_0$

We can reformulate expression (19) as follows

$$\log \mathbf{L}_0 = \#F \cdot \log(2M) + \sum_{f \in F} \underbrace{\left[ \underbrace{\log |\hat{A}_n^f|}_{L_f} - M \underbrace{|\hat{A}_n^f|^2}_{B_f} \right]}_{S_f} \quad (24)$$

The r.v.'s  $|\hat{a}_n^f|$  are i.i.d. (assuming independency among DCT frequencies) and they follow a Rayleigh distribution with scale parameter  $\sigma = 1/\sqrt{2M}$ . Starting from this we can study  $L_f$ ,  $B_f$  and  $S_f$ .

- Each  $L_f$  is a log-Rayleigh random variable. According to [34], we have that

$$E\{L_f\} = -\frac{\log M}{2} - \frac{\gamma}{2}, \quad Var\{L_f\} = \frac{\pi^2}{24},$$

where  $\gamma$  is the Euler-Mascheroni constant.

- Each  $B_f$  is a squared Rayleigh r.v. multiplied by a constant term  $-M$ . It can be shown (via r.v. transformation) that a squared Rayleigh r.v. with scale parameter  $\sigma$  has an exponential distribution with rate parameter  $1/2\sigma^2$  (in our case  $M$ ). By scaling with a factor  $-M$ , we have that

$$E\{B_f\} = -1, \quad Var\{B_f\} = 1.$$

- Each  $S_f$ 's is sum of two r.v., hence we have that

$$E\{S_f\} = E\{L_f\} + E\{B_f\} = -\frac{\log M}{2} - \frac{\gamma}{2} - 1,$$

$$Var\{S_f\} = Var\{L_f\} + Var\{B_f\} + 2Cov\{L_f, B_f\} = \frac{\pi^2}{24}$$

as the value of the covariance has been derived by means of symbolic computation and is equal to  $-1/2$ .

- Finally,  $\log \mathbf{L}_0$  is a sum of the iid r.v.'s  $S_f$  plus a constant term  $\#F \cdot \log(2M)$ . Then, we have that

$$E\{\log \mathbf{L}_0\} = \#F \cdot \log(2M) - \#F \cdot \left( \frac{\log M}{2} + \frac{\gamma}{2} + 1 \right),$$

$$Var\{\log \mathbf{L}_0\} = \#F \cdot \frac{\pi^2}{24}.$$

## REFERENCES

- [1] H. Sencar and N. Memon, Eds., *Digital Image Forensics - There is more to a picture than meets the eye*. Springer, 2013.
- [2] A. Rocha, W. Scheirer, T. Boult, and S. Goldenstein, "Vision of the unseen: current trends and challenges in digital image and video forensics," *ACM Computing Surveys*, vol. 43, no. 4, 2011.
- [3] M. Stamm, M. Wu, and K. Liu, "Information forensics: an overview of the first decade," *IEEE Access*, pp. 167–200, 2013.
- [4] I. Steadman. (2013, May) "Fake" World Press Photo isn't fake, is lesson in need for forensic restraint. [Online]. Available: <http://www.wired.co.uk/news/archive/2013-05/16/photo-faking-controversy>
- [5] G. Scoble. (2015, June) How World Press Photo catches image manipulators. [Online]. Available: <http://www.pdnonline.com/news/How-World-Press-Photo-Catches-Image-Manipulators-13819.shtml>
- [6] F. Pérez-González, T. Quach, S. J. Miller, C. Abdallah, and G. Heileman, "Application of Benford's law to images," S. J. Miller, A. Berger and T. Hill (Eds), *The Theory and Applications of Benford's law*, Princeton University Press, 2015.
- [7] C. Pasquini, F. Pérez-González, and G. Boato, "A Benford-Fourier JPEG compression detector," in *Proc. of IEEE ICIP*, 2014, pp. 5322–5326.
- [8] Z. Fan and R. D. Queiroz, "Identification of bitmap compression history: JPEG detection and quantizer estimation," *IEEE Transactions on Image Processing*, vol. 12, n. 2, pp. 230–235, 2003.
- [9] R. Neelamani, R. D. Queiroz, Z. Fan, S. Dash, and R. Baraniuk, "JPEG compression history estimation for color images," *IEEE Transactions on Image Processing*, vol. 15, n. 2, pp. 1363–1378, 2006.
- [10] W. Luo, J. Huang, and G. Qiu, "JPEG error analysis and its applications to digital image forensics," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 3, pp. 480–491, 2010.
- [11] B. Li, T.-T. Ng, X. Li, S. Tan, and J. Huang, "Revealing the trace of high-quality JPEG compression through quantization noise analysis," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 3, pp. 558–573, 2015.
- [12] J. Yang, G. Zhu, J. Huang, and X. Zhao, "Estimating JPEG compression history of bitmaps based on factor histogram," *Digital Signal Processing*, no. 41, pp. 90–97, 2015.
- [13] D. Fu, Y. Shi, and W. Su, "A generalized Benford's law for JPEG coefficients and its applications in image forensics," in *SPIE Conference on Security, Steganography, and Watermarking of Multimedia Contents*, vol. 6505, 2007.
- [14] F. Pérez-González, G. Heileman, and C. Abdallah, "Benford's law in image processing," in *Proc. of IEEE ICIP*, 2007, pp. 405–408.
- [15] C. Pasquini and G. Boato, "JPEG compression anti-forensics based on first significant digit distribution," in *Proc. of IEEE MMSP*, 2013, pp. 500–505.
- [16] P. Comesaña-Alfaro and F. Pérez-González, "The optimal attack to histogram-based forensic detectors is simple(x)," in *Proc. of IEEE WIFS*, 2014, pp. 1730–1735.
- [17] T. Birney and T. Fischer, "On the modeling of DCT and subband image data for compression," *IEEE Transactions on Image Processing*, vol. 4, no. 2, pp. 186–193, 1995.
- [18] J. Barry, E. Lee, and D. Messerschmitt, "Digital communication," Springer, 2004.
- [19] S. Rice, "Mathematical analysis of random noise," *Bell System Technical Journal*, vol. 24, pp. 46–156, 1945.
- [20] T. Bianchi and A. Piva, "Image forgery localization via block-grained analysis of JPEG artifacts," *IEEE Transactions on Information Forensics and Security*, vol. 7, n. 3, pp. 1003–1017, 2012.
- [21] C. Pasquini, G. Boato, and F. Pérez-González, "Multiple JPEG compression detection by means of Benford-Fourier coefficients," in *Proc. of IEEE WIFS*, 2014, pp. 113–118.
- [22] J. Havil, *Gamma: exploring Euler's constant*, ser. Princeton Science Library. Princeton University Press, 2009.
- [23] Q. Liu, "Detection of misaligned cropping and recompression with the same quantization matrix and relevant forgery," in *ACM Workshop on Multimedia Forensics and Intelligence*, 2011, pp. 25–30.
- [24] G. Schaefer and M. Stich, "UCID - An uncompressed colour image database," in *SPIE Storage and Retrieval Methods and Applications to Multimedia*, vol. 5307, 2004.
- [25] T. Gloe and R. Boehme, "The Dresden image database for benchmarking digital image forensics," in *ACM Symposium on Applied Computing*, vol. 2, 2010, pp. 1585–1591.
- [26] D.-T. Dang-Nguyen, C. Pasquini, V. Conotter, and G. Boato, "RAISE - a raw images dataset for digital image forensics," in *Proc. of ACM MMSys*, 2015, pp. 219–224.
- [27] C. Pasquini, P. Schöttle, R. Böhme, G. Boato, and F. Pérez-González, "Forensics of high quality and nearly identical JPEG image recompression," in *ACM Information Hiding & Multimedia Security (IH&MMSec)*, 2016, pp. 11–21.
- [28] M. Stamm and K. J. Ray Liu, "Anti-forensics of digital image compression," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 3, pp. 1050–1065, 2011.
- [29] M. Kirchner and S. Chakraborty, "A second look at first significant digit histogram restoration," in *Proc. of IEEE WIFS*, 2015, pp. 1–6.
- [30] A. De Rosa, M. Fontani, M. Massai, A. Piva, and M. Barni, "Second-order statistics analysis to cope with contrast enhancement counter-forensics," *IEEE Signal Processing Letters*, vol. 22, no. 8, pp. 1132–1136, 2015.
- [31] T. Bianchi and A. Piva, "Reverse engineering of double JPEG compression in the presence of image resizing," in *Proc. of IEEE WIFS*, 2012, pp. 127–132.
- [32] Q. Liu, "An approach to detecting JPEG down-recompression and seam carving forgery under recompression anti-forensics," *Pattern Recognition*, vol. 65, pp. 35–46, 2016.
- [33] H. Kobayashi, B. Mark, and W. Turin, *Probability, Random Processes and Statistical Analysis*. Cambridge University Press, 1991.
- [34] B. Rivet, L. Girin, and C. Jutten, "Log-Rayleigh distribution: a simple and efficient statistical representation of log-spectral coefficients," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 3, 2007.



**Cecilia Pasquini** is a Post-Doctoral Researcher at the Privacy and Security Lab with the Universität Innsbruck, Austria, and University of Münster, Germany. She received the Ph.D. degree from the University of Trento, Italy, in 2016. She has been a Visiting Scholar with the University of Vigo, Spain, in 2013. Her research interests include multimedia forensics and anti-forensics, multimedia analysis and verification.



**Giulia Boato** is an Assistant Professor with the Department of Information Engineering and Computer Science, University of Trento, Italy, where she has been coordinating the Multimedia Signal Processing and Understanding Lab from 2008 to 2011. She was on the project staff of many projects (FP7 FET-IP LIVINGKNOWLEDGE, FP7 IP GLOBAL, FP7 CA ETERNALS, and ICT-COST 3D-ConTourNet). She has co-authored 85 papers in international conferences and journals. Her research interests include image and signal processing, with particular attention to multimedia data protection, data hiding and digital forensics but also intelligent multidimensional data management and analysis.



**Fernando Pérez-González** (M'90–SM'09–F'16) is a Professor with the School of Telecommunication Engineering, University of Vigo, Spain, and also a Research Professor in information science and technology with the University of New Mexico, USA. From 2007 to 2010, he was Program Manager of the Spanish National Research and Development Plan on Electronic and Communication Technologies, Ministry of Science and Innovation. From 2007 to 2014, he was the Executive Director of the Galician Research and Development Center in Advanced Telecommunications. He has co-authored over 60 papers in international journals, 160 conference papers and several international patents. His research interests include the areas of digital communications, adaptive algorithms, privacy enhancing technologies, and information forensics and security.