A BENFORD-FOURIER JPEG COMPRESSION DETECTOR

Cecilia Pasquini [*], *Fernando Pérez-González* [†,‡], *Giulia Boato* [*]

[*] DISI, University of Trento
38123 Trento-ITALY
cecilia.pasquini@unitn.it, boato@disi.unitn.it
[†] EE Telecomunicacion, University of Vigo
[‡] GRADIANT (Galician Research and Development Center in Advanced Telecommunications)
36310 Vigo-SPAIN
fperez@gts.uvigo.es

ABSTRACT

Intrinsic statistical properties of natural uncompressed images can be used in image forensics for detecting traces of previous processing operations. In this paper, we extend the recent theoretical analysis of Benford-Fourier coefficients and propose a novel forensic detector of JPEG compression traces in images stored in an uncompressed format. The classification is based on a binary hypothesis test for which we can derive theoretically the confidence intervals, thus avoiding any training phase. Experiments on real images and comparisons with state-of-art techniques show that the proposed detector outperforms existing ones and overcomes issues due to dataset-dependency.

Index Terms— Binary hypothesis testing, Image Forensics, JPEG Compression.

1. INTRODUCTION

Multimedia forensics techniques deal with the recovery of information that can be used to authenticate and measure the trustworthiness of digital multimedia contents. The inspiring principle is that inherent traces are left in a digital data both when the content is created and during successive processing.

In this framework, the traces left by JPEG compression, the most popular coding scheme for images, have been widely studied and exploited in different forensic scenarios, allowing the identification of previous compression in uncompressed images [1, 2, 3], of the number of compression steps applied to a given content (e.g., detecting a double compression) [4, 5, 6], of inconsistencies of the traces within the same image [7], and possibly of the coding parameters used [8]. In an adversarial perspective, procedures designed to hide these kind of traces have also been proposed, which modify the DCT coefficients [9, 10, 11] and their First Significant Digits [12, 13, 14].

In this paper we focus on the problem of detecting JPEG compression traces in images stored in uncompressed format. A first approach to this issue was presented in [1] with an analysis of blocking artifacts left by the 8×8-block DCT transformation within the JPEG compression, and then further improved in [8]. The distribution of First Significant Digits (FSD) of DCT coefficients is exploited in [2], since for uncompressed images it can be modeled very accurately by Benford's law, while the compression introduces a deviation from this known distribution or a generalized version of it. Finally, in [3] a theoretic study about the distribution of DCT coefficients in uncompressed images is carried out, where the Benford-Fourier coefficients are defined in the first place and, in [15], exploited as features to train an SVM, achieving very high accuracies.

Inspired by the results obtained in [3] and [15], in this paper we extend the Benford-Fourier theory and design a binary decision test to distinguish natural uncompressed images from images that underwent a generic processing, in particular JPEG compression. The statistics used are an estimate of the Benford-Fourier coefficients obtained from the analyzed image by using a sample mean, thus avoiding the computation of any histogram (which can generate ambiguities due to the choice of the bin width) and leading to a very low computational complexity. Moreover, we have derived the pdf of such estimate for the case of uncompressed images as a function of the size of the image, which allows us to theoretically determine a threshold once that an upper bound to the false alarm probability is fixed. As a result, the proposed detector is size-adaptive and does not require a training phase to be performed on an extensive data collection, as usually happens in state-of-art statistically-based techniques.

In the following, the existing Benford-Fourier theory and the novel extensions are presented in Section 2, while in Section 3 we report the results of the experiments we performed with different datasets and detectors. Finally, in Section 4 we conclude the paper discussing open problems and future research directions.

2. BENFORD-FOURIER COEFFICIENTS FOR FORENSIC ANALYSIS

In this section, we present the Benford-Fourier theory, on which our forensic analysis is based. We first review the results reported in [3], which constitute the basis of our new study on the statistical estimation of Benford-Fourier coefficients starting from a given image, that is presented in Section 2.2. Finally, we design the binary hypothesis test used to classify images as uncompressed or JPEG compressed.

2.1. Statistics of Benford-Fourier coefficients

In [3], the authors have studied the distribution of DCT coefficients in the modular logarithmic domain. Precisely, they considered a random variable X with a symmetric probability density function $f_X(x)$ with respect to 0. The following theoretical analysis holds both when X models *all* the 8×8-block DCT coefficients of a 2-D

image and when only the coefficients of given frequency are considered. In this paper, we exploit the second case, i.e., the r.v. we study is related to a generic frequency (i, j), although we will omit this dependency for the sake of clarity.

Then, Z is the r.v. determined by the pdf

$$f_Z(z) = \begin{cases} f_Z(z) = 2f_X(z), & z > 0 \\ 0, & z \leq 0. \end{cases}$$

(thus modeling with Z the behavior of $|X|$ in $\mathbb{R}_0^+$) and the distribution of the following r.v.'s is studied:

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

$$\tilde{Z} = \log_{10} Z \mod 1.$$

It has been shown that, for a generic continuous r.v. Z, the pdf of $\tilde{Z}$ can be expressed as follows:

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

where $a_n = |a_n|e^{j\phi_n}$ is the Fourier transform of $f_{Z'}(z')$ evaluated at $2\pi n$, with $n \in \mathbb{N}$. The values a_n are called *Benford-Fourier coefficients (BF)*.

As it is has been widely investigated in the literature [16], in natural uncompressed images the distribution of X is accurately modeled by a Generalized Gaussian (GG) pdf with standard deviation σ and shaping factor c:

$$f_X(x) = \frac{\beta c}{2\Gamma(1/c)} e^{-|\beta x|^c}, \quad \beta = \frac{1}{\sigma}\sqrt{\frac{\Gamma(3/c)}{\Gamma(1/c)}}.$$

For such case, the expression of a_n, $n \in \mathbb{N}$, and their magnitude have been derived in [3]

$$a_n = \frac{2A}{\beta c} e^{j\frac{2\pi n \log\beta}{\log 10}} \Gamma\Big(\frac{-j2\pi n + \log 10}{c \log 10}\Big) \tag{1}$$

$$|a_n| = \sqrt{\prod_{k=0}^{\infty} \left[1 + \frac{(2\pi n)^2}{\log^2 10(ck+1)^2}\right]^{-1}}. \tag{2}$$

As seen in (2), the magnitude of the coefficients increases with c and does not depend on the variance of the GG; in particular, when $n \geq 3$ and $c \leq 2$ (which is the case of natural uncompressed images), these values are lower than 10^{-2}. This suggests that the behavior of the $a_n, n \in \mathbb{N}$ can be used to characterize uncompressed images and detect images that underwent some processing. In [15], a first approach in this direction has been proposed, which discriminates uncompressed and JPEG compressed images.

2.2. Estimation of Benford-Fourier coefficients

In order to use BF coefficients for analyzing an image and determine whether it is uncompressed or not, we need a numerical procedure to estimate such coefficients given the subject image.

By definition, the BF coefficients are given by

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

In other words, they are 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.

We can then obtain an estimation of such coefficients starting from a set of M samples by computing the sample mean

$$\hat{a}_n = \frac{\sum_{m=1}^{M} e^{-j2\pi n \log_{10} Z_m}}{M} \qquad m = 1, \ldots, M \tag{3}$$

where Z_m are the r.v.'s representing the DCT coefficients at the chosen frequency in the m-th block, that we suppose independent and identically distributed.

Although the sample mean is a minimum variance unbiased estimator of the expected value (i.e., $E\{\hat{a}_n\} = a_n$), we should consider the fact that the actual accuracy of $\hat{a}_n$ as estimator of a_n depends on the size of the sample considered. For this reason, we are interested in studying its distribution 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 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 [17]. In other words,

$$\hat{a}_n = a_n + W_0$$

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 [17]) is that $E\{W_0^2\} = 0$. Starting from the definition of $\hat{a}_n$, it is easy to prove that

$$E\{W_0^2\} = E\{(\hat{a}_n - a_n)^2\} = \frac{1}{M}(a_{2n} - a_n^2). \tag{4}$$

As we mentioned before, $|a_n|$ is very small for uncompressed images and the value of (4) will be very close to 0, therefore $\hat{a}_n$ is approximately a circular bivariate normal r.v. with non-zero mean.

It is well known that the distribution of $R := |\hat{a}_n|$ is given by the Rice distribution with parameters $|a_n|$ and σ, where σ is the standard deviation of its real and imaginary parts [18]. Similarly as before, we can now obtain σ^2 by exploiting the fact that for a Rice distribution

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

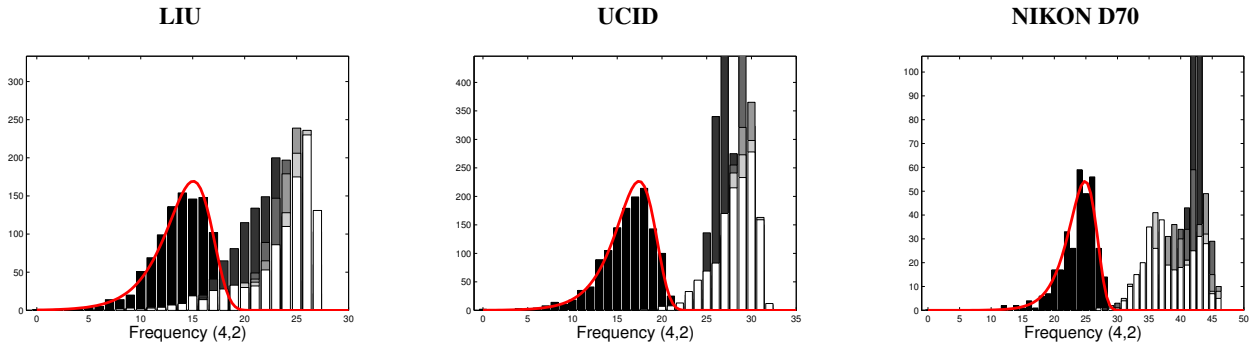
As we observed, $|a_n|$ is extremely small when $n \geq 3$. This leads to serious issues in the evaluation of the Rice pdf, due to numerical overflow/underflow of the exponential term and the Bessel function, respectively. Because of that, we can reasonably assume $|a_n| \approx 0$ and consider the special case of Rice distribution with mean parameter 0, i.e., the Rayleigh distribution, and $\sigma = 1/\sqrt{2M}$:

$$f_R(r) = 2Mre^{-Mr^2}. \tag{5}$$

This formula allows us to predict the distribution of BF coefficients $\hat{a}_n$ estimated from an uncompressed natural image by means of (3). From the properties of the Rayleigh distribution, we have that

$$\mu_R = \frac{1}{\sqrt{M}} \cdot \frac{\sqrt{\pi}}{2}, \quad \sigma_R^2 = \frac{1}{M} \cdot \frac{4 - \pi}{4}.$$

It means that R is in any case an overestimate whose expected accuracy increases linearly with $\sqrt{M}$ and its variance decreases linearly with M.



$E_{(4,2)} = 3.4 \cdot 10^{-2}$ $E_{av} = 2.9 \cdot 10^{-2}$

$E_{(4,2)} = 3.4 \cdot 10^{-2}$ $E_{av} = 2.6 \cdot 10^{-2}$

$E_{(4,2)} = 6.9 \cdot 10^{-2}$ $E_{av} = 5.7 \cdot 10^{-2}$

Fig. 1. Histograms of the BF coefficients of the frequency $(4, 2)$ estimated from each dataset with $n = 4$ are reported in grayscale according to their quality factors. Because of visual issues, the x-axis has been scaled by a factor of M (equal to the number of blocks in the images of each dataset) and converted to dB. Squared sum differences between the theoretical model (the Rayleigh pdf in red) and the data obtained are also reported.

2.3. Design of a binary hypothesis test

Since we know the pdf of the BF coefficients for the case of uncompressed images, we can now design a binary hypothesis test on the statistics $R = |\hat{a}_n|$ (for some $n \geq 3$), where

$$H_0\text{: the image is uncompressed}$$

$$H_1\text{: the image is single compressed}$$

The upper and lower thresholds U and L (the test is two-tailed) can be theoretically derived by fixing a type I error probability and using the cdf of the Rayleigh distribution in (5), given by

$$F_R(r) = 1 - e^{-Mr^2}.$$

As a general approach, once a type I error probability upper bound α is fixed, we can reject the null hypothesis when the value of r obtained from the image is such that

$$r \leq L \quad \text{or} \quad r \geq U$$

where

$$F_R(L) = \frac{\alpha}{2} \quad \text{and} \quad 1 - F_R(U) = \frac{\alpha}{2}.$$

3. EXPERIMENTAL RESULTS

In order to verify the validity of our theoretic predictions, we carried out several experimental sessions on three datasets of uncompressed color images with different sizes, each of them being previously used in the literature for image forensics:

- **LIU:** a set of 1000 uncompressed images (resolution 256×256, 1024 blocks) selected from the database used in [19];
- **UCID:** a set of 1338 uncompressed images (resolution 384×512, 3072 blocks) from the UCID [20];
- **NIKON D70:** a set of 320 uncompressed images (resolution 2000×3008, 94000 blocks) selected from the Dresden database [21].

For each dataset, we first computed the BF coefficients at every DCT frequency on the luminance channel for the uncompressed images, and then for the same images compressed at quality factors $\{90, 80, 70, 60, 50\}$. Matlab built-in functions have been used for applying this processing. We report their histograms in Figure 1, where the uncompressed case is represented by black bars, and quality factors $90, 80, 70, 60, 50$, respectively, are represented by grayscale tones fading to white. In red, we plot the Rayleigh distribution, computed as in (5). For each dataset, we also report the squared sum of the differences between the predicted probabilities and the empirical ones, both for the frequency considered and the average value over all the 64 frequencies. As it can be seen, in each case the actual distribution of the BF coefficients for uncompressed images is accurately modeled by the Rayleigh pdf predicted from the theory. In this figure, we considered $n = 4$ (so that $|a_n|$ is very small and the approximation $|a_n| \approx 0$ holds) and the frequency $(4, 2)$, which are the values that we used for our experiments, but similar results were obtained for different frequencies and values of $n \geq 2$. On the other hand, the behavior of BF coefficients for compressed images depends on the quality factor and the frequency considered. We empirically chose the frequency $(4, 2)$ from preliminary tests as the one that achieved the higher average precision over the three datasets with an average false alarm rate lower than 0.01.

We compared our detector with type I error probability upper bound equal to 0.01 (**`NEW BF`**), with three methods proposed in the literature for the same forensic problem. The first one is the approach proposed in [15] (indicated as **`BF SVM`**), which is also based on BF theory but there the authors consider the DCT coefficients of the whole image and estimate the BF coefficients by computing the FFT of the empirical distribution of $\tilde{Z}$. Then, the first five coefficients are used as feature to train an SVM discriminating between natural uncompressed images and images that underwent a JPEG compression. The second one, **`FSD SVM`**, was proposed in [2] and extracts the empirical frequencies of the 9 significant digits of DCT coefficients as features for training an SVM. Finally, we considered the compression detector presented in [1] (**`BLOCK`**), where inter- and intra-block pixel differences are computed and combined in a final statistic K, expressing the strength of blocking artifacts in the pixel domain: images presenting a value of K higher than a certain threshold are classified as compressed.

		LIU						UCID						NIKON D70					
		90	**80**	**70**	**60**	**50**	**R**	**90**	**80**	**70**	**60**	**50**	**R**	**90**	**80**	**70**	**60**	**50**	**R**
NEW BF	**FA**				0.4						1.2						0.0		
	DP	85.6	94.8	92.8	89.2	86.0	94.0	99.1	99.7	99.4	98.8	98.2	99.4	100	100	100	100	100	100
	ACC	92.6	97.2	96.2	94.4	92.8	96.8	99.0	99.3	99.1	98.8	98.5	99.1	100	100	100	100	100	100
BF SVM	**FA**				23.2						5.24						3.1		
	DP	99.6	99.8	97.4	98.6	98.0	98.6	99.5	99.4	97.8	99.4	99.1	98.8	100	99.4	93.2	100	100	98.8
	ACC	88.2	88.3	87.1	87.7	87.4	87.7	97.2	97.1	96.3	97.1	97.0	96.8	98.4	98.2	95.0	98.5	98.5	97.8
FSD SVM	**FA**				14.0						2.6						0.0		
	DP	71.4	99.6	95.2	79.6	72.4	96.0	89.2	99.6	99.4	96.7	96.4	98.6	98.2	100	93.7	68.1	57.5	97.5
	ACC	78.7	92.8	90.6	82.8	79.2	91.0	93.3	98.5	98.4	97.1	96.9	98.1	99.0	100	96.8	84.0	78.7	98.7
BLOCK	**FA**				0.8						1.2						0.0		
	DP	53.2	71.4	80.8	88.0	91.6	71.6	94.9	99.8	100	100	100	98.9	100	100	100	100	100	100
	ACC	76.2	85.3	90.0	93.6	95.4	85.4	96.8	99.3	99.4	99.4	99.4	98.8	100	100	100	100	100	100

Table 1. The table contains the results obtained by applying the four methods considered to the three different datasets. In each subtable, we report the performance of the detector in terms of false alarm rate, detection rate, and accuracy percentage. In the first five columns, the testing set is composed of uncompressed images and images compressed with a fixed quality factor in $\{50, 60, 70, 80, 90\}$ (the same as the training set), while in the last column (labeled as **R**) we considered images compressed with random quality factors in $\{50, 51, \ldots, 89, 90\}$.

These previous methods require a dataset for a training phase: training features need to be available for the SVMs and in the **BLOCK** method it is necessary to determine a good threshold value for the classification. Therefore, each dataset was randomly divided in two equal parts, one for training, and another for testing. Images in the former were compressed at quality factors $\{90, 80, 70, 60, 50\}$, features were extracted from each version (as well as the original uncompressed ones) and used for training the SVMs or determining the optimal threshold. Finally, we created a first testing set by using the same quality factors as in the training set and a second one with random quality factors in $\{50, 51, \ldots 89, 90\}$.

We report the results obtained by applying each method on the testing set in Table 1. It can be observed that the two methods based on an SVM behave quite differently when changing the dataset and the quality factor, especially in terms of false alarm. On the other hand, the performance of **NEW BF** and **BLOCK** are very good for **UCID** and **NIKON D70** datasets, while for the **LIU** dataset the method **BLOCK** shows a generally low detection probability.[1]

In general, determining the threshold according to the size of the image helps our method in keeping a good detection rate, with a low false alarm rate. Moreover, we should consider that for the other three methods the training phase has been carried out separately for each dataset and could lead to different results when applied to an image with different size and source. For instance, by applying the method **BLOCK** (which achieves very good performance in both cases) on **UCID** using the threshold trained from **NIKON D70**, we obtain a very high false alarm rate. This is due to the fact that, according to our experimental results, the values of K for uncompressed images generally decrease as the image size increases, but no theoretical model that makes it possible to generalize the classifier is available for such statistics.

[1]The method **BF SVM** was originally applied to 8-bit grayscale images, while here we work on the luminance channel in a non-integer domain. For this reason, together with the random choice of the training and testing sets, the results on the **UCID** dataset are slightly different with respect to those reported in [15], where the SVM classifier achieved better performance.

4. CONCLUSIONS

We have proposed a method that detects JPEG compression operations in images stored in uncompressed formats, which is based on the extraction of a single feature. The results of our experiments on three different datasets and the comparison with state-of-the-art methods are encouraging and show that the size-adaptive hypothesis test helps in increasing the flexibility of the detector with respect to the images considered, despite the low computational complexity and the absence of a training phase.

However, some issues are still open and will be subject of future research. The first one is that currently the choice of the frequency used is based on preliminary tests, although it is constant over the different datasets. Hence, the development of an effective strategy to combine the estimation of BF coefficients obtained from multiple and automatically selected frequencies is a primary goal. To this end, it is worth pointing out that the frequency selection process strictly depends on the effects of the processing operation considered, in this case the JPEG compression, since the model for natural uncompressed images is accurate for any frequency. Indeed, deriving a theoretical model also for the alternative hypothesis would be another step forward, since it would completely determine the test and open the way for the estimation of the quantization table. Finally, we remark that the binary hypothesis test could be adapted to the detection of other processing operations that modify the distribution of the DCT coefficients, since the theoretical model applies to the null hypothesis. For instance, studying the effects of double or multiple JPEG compression seems a natural extension of this work.

5. REFERENCES

[1] Z. Fan and R. De Queiroz, "Identification of bitmap compression history: JPEG detection and quantizer estimation," *IEEE Transactions on Image Processing*, vol. 12, n. 2, pp. 230–235, 2003.

[2] D. Fu, Y.Q. Shi, and W. Su, "A generalized Benford's law for JPEG coefficients and its applications in image forensics," in *Proceedings of SPIE Conference on Security, Steganography, and Watermarking of Multimedia Contents*, 2007, vol. 6505.

[3] F. Pérez-González, G.L. Heileman, and C.T. Abdallah, "Benford's law in image processing," in *Proceedings of IEEE ICIP*, 2007, pp. 405–408.

[4] 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.

[5] S. Milani, M. Tagliasacchi, and S. Tubaro, "Discriminating multiple JPEG compression using first digit features," in *Proceedings of IEEE ICASSP*, 2012, pp. 25–30.

[6] Y.L. Chen and C.T. Hsu, "Detecting recompression of JPEG images via periodicity analysis of compression artifacts for tampering detection," *IEEE Transactions on Information Forensics and Security*, vol. 6, n. 2, pp. 396–406, 2011.

[7] M. Fontani, T. Bianchi, A. De Rosa, A. Piva, and M. Barni, "A framework for decision fusion in image forensics based on Dempster-Shafer theory of evidence," *IEEE Transactions on Information Forensics and Security*, vol. 8, n. 4, pp. 593–607, 2013.

[8] R. Neelamani, R. De Queiroz, Z. Fan, S. Dash, and R.G. Baraniuk, "JPEG compression history estimation for color images," *IEEE Transactions on Image Processing*, vol. 15, n. 2, pp. 1363–1378, 2006.

[9] M.C. 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.

[10] M. Barni, M. Fontani, and B. Tondi, "A universal technique to hide traces of histogram-based image manipulations," in *Proceedings of ACM Workshop on Multimedia and Security*, 2012, pp. 97–104.

[11] P. Comesaña-Alfaro and F. Pérez-González, "Optimal counterforensics for histogram-based forensics," in *Proceedings of IEEE ICASSP*, 2013, pp. 3048–3052.

[12] S. Milani, M. Tagliasacchi, and S. Tubaro, "Antiforensics attack to Benford's law for the detection of double compressed images," in *Proceedings of IEEE ICASSP*, 2013, pp. 3053–3057.

[13] C. Pasquini and G. Boato, "JPEG compression anti-forensics based on first significant digit distribution," in *Proceedings of IEEE MMSP*, 2013, pp. 500–505.

[14] C. Pasquini, P. Comesaña-Alfaro, F. Pérez-González, and G. Boato, "Transportation-theoretic image counterforensics to First Significant Digit histogram forensics," in *Proceedings of ICASSP*, 2014, pp. 2718–2722.

[15] F. Pérez-González, T.T. Quach, S. J. Miller, C.T. Abdallah, and G.L. Heileman, "Application of Benford's law to images," *S. J. Miller, A. Berger and T. Hill (Eds), Theory and Applications of Benford's law, Princeton University Press*, 2014.

[16] F. Muller, "Distribution shape of two-dimensional DCT coefficients of natural images," *IEEE Electronic letters*, vol. 29, n. 22, pp. 1935–1936, 1993.

[17] J.R. Barry, E.A. Lee, and D.G. Messerschmitt, "Digital communication," *Springer*, 2004.

[18] S.O. Rice, "Mathematical analysis of random noise," *Bell System Technical Journal*, vol. 24, pp. 46–156, 1945.

[19] Q. Liu, "Detection of misaligned cropping and recompression with the same quantization matrix and relevant forgery," in *Proceedings of ACM Workshop on Multimedia Forensics and Intelligence*, 2011, pp. 25–30.

[20] G. Schaefer and M. Stich, "UCID - An uncompressed colour image database," in *Proceedings of SPIE Storage and Retrieval Methods and Applications to Multimedia*, 2004, vol. 5307.

[21] T. Gloe and R. Boehme, "The Dresden image database for benchmarking digital image forensics," in *Proceedings of ACM Symposium on Applied Computing*, 2010, vol. 2, pp. 1585–1591.