TOWARDS A THEORY OF JPEG BLOCK CONVERGENCE

Cecilia Pasquini and Rainer Böhme

Department of Computer Science, Universität Innsbruck, 6020 Innsbruck, Austria
Department of Information Systems, University of Münster, 48149 Münster, Germany

ABSTRACT

The convergence statistics of JPEG blocks has been shown to be a useful tool to forensically analyze high quality compressed images. Since current approaches are based on empirical observations, we propose a theoretical analysis explaining the case of grayscale images and maximum quality JPEG compression (i.e., quality factor equal to 100). The approximate distribution of the stable block ratio at different compression stages is derived, showing that it ultimately depends on the variance of the quantization noise in the DCT domain. We apply such results to discriminate never compressed images and images compressed once with maximum quality, by resorting to results on JPEG error statistics. Tests on image patches with different size and content validate the theoretical results, which allow for obtaining high accuracy through a calibration-free maximum likelihood classification rule.

Index Terms— Image forensics, high quality compression, block convergence, hypothesis test, statistical modeling.

1. INTRODUCTION

The traces of JPEG compression in digital images have been widely investigated for forensic purposes, by studying the effect of both the partition of the spatial domain in 8×8 blocks and the quantization in the DCT domain. It is known that high quality compression, possibly repeated with identical parameters, is a challenging case. The statistical traces left by such operations are generally weak and histogram-based techniques lose accuracy. Thus, specific approaches have been developed for the detection of high quality compression [1] and identical recompression [2] (i.e., with the same quantization table). Special attention has been devoted to the maximum quality case [3, 4], i.e., when the quality factor is equal to 100 and the quantization step used in the DCT domain is 1 for all subbands. For brevity, we refer to this kind of compression as JPEG-100.

The detectors based on JPEG block convergence are valuable tools because of their minimal dependence on the image content and their ability to accurately identify several previous compressions (up to 10). The main rationale is the fact that the 8×8 image blocks (processed independently during the JPEG compression scheme) will at some point get *stable* (i.e., completely unaltered) in the spatial domain after repeated identical recompression. This process, called *block convergence*, occurs in a similar manner for all blocks. Thus, by recompressing with JPEG-100 the image under investigation and computing the ratio R of blocks that turn out to be stable, it is possible to estimate how many previous JPEG-100 compressions occurred. This has been first shown in [3] for grayscale images, successively extended to color images and other high quality factors (≥ 90) in [5], and combined with DCT-based approaches in [6].

This approach is currently based on the empirical observation of this property. In fact, a calibration phase, where R is computed on differently processed images and its empirical distribution is fitted, is always necessary. However, we lack a theoretical understanding of why these ratios are similarly obtained across different image content and size. In this work, we address this gap for the case of grayscale images and JPEG-100 compression, by deriving a model for the pdf of R. We obtain that its mean mainly depends on the number of previous JPEG-100 compressions, while its variance is related to the number of blocks involved in the analysis. Building on recent results on JPEG error statistics [7], we apply our theory to the problem of discriminating never compressed from decompressed images. The decision is made according to the Maximum Likelihood rule among the two theoretically derived distributions of R, thus no calibration is necessary. Extensive tests show that the obtained accuracy is high even for small image patches.

The paper is structured as follows: Section 2 recalls the JPEG-100 compression pipeline and introduces the notations used throughout the paper; in Section 3, we develop the theoretical model of the stable block ratio distribution; we instantiate the model for the case of never compressed versus decompressed images by integrating existing results in Section 4; experimental tests on real images are proposed in Section 5, and conclusions are drawn in Section 6.

2. NOTATION AND COMPRESSION SCHEME

To recall the processing steps carried out on an image block during JPEG-100 compression, we define $\mathbf{X}_t$ as the random vector whose components $\mathbf{X}_t[i]$, $i = 1,\dots,64$, represent the integer pixel values of a generic 8×8 block (arranged column-wise) in a grayscale image compressed t times with JPEG-100. We denote the direct and inverse two-dimensional DCT by the operators $\mathrm{DCT2}(\cdot)$ and $\mathrm{IDCT2}(\cdot)$, and we define $\mathbf{Y}_t \doteq \mathrm{DCT2}(\mathbf{X}_t)$. Most approaches neglect the finite arithmetic errors introduced by these computations (i.e., it is assumed that $\mathrm{IDCT2}(\mathrm{DCT2}(\mathbf{X}_t)) = \mathbf{X}_t$). We provisionally adopt this assumption and will discuss its validity in Section 5.

In the case of JPEG-100 compression, the DCT coefficients of the image block are quantized with step 1. During decompression, the inverse DCT is applied to the quantized values, and the resulting pixel values are rounded to the nearest integer in the range defined by the bit depth. As the operations in the spatial and DCT domain are essentially equivalent (except for saturated pixels), we can define the operator $[\cdot]$, which rounds to the nearest integer every component of its argument, and represent the compression pipeline as in Fig. 1. Two auxiliary random vectors

$$\tilde{\mathbf{Y}}_t \doteq [\mathbf{Y}_t], \quad \tilde{\mathbf{X}}_t \doteq \mathrm{IDCT2}(\tilde{\mathbf{Y}}_t) \tag{1}$$

are defined for convenience.
After rounding $\tilde{\mathbf{X}}_t$, we obtain the final block $\mathbf{X}_{t+1}$, which is the starting point for a potential subsequent recompression.

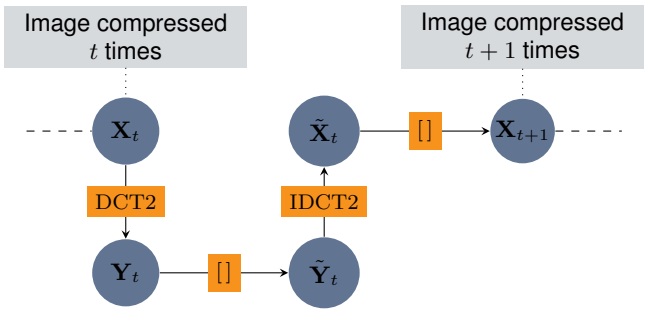


Fig. 1. JPEG-100 compression pipeline.

3. PROPOSED MODEL

As mentioned in Section 1, forensic methods based on block convergence analyze the ratio of so-called *stable* blocks, i.e., blocks that are completely unaltered in the spatial domain after recompression with JPEG-100. In particular, it has been shown experimentally that the more JPEG-100 compression stages have previously occurred, the more the ratio of stable blocks increases towards 1. With the goal of obtaining a statistical characterization of this phenomenon, we derive a model for the following quantities:

- the probability p_t that a single block is stable in an image compressed t times with JPEG-100 (Section 3.1);
- the distribution of the stable block ratio R among N blocks of an image compressed t times with JPEG-100 (Section 3.2).

3.1. Single block statistics

Using the notation of Section 2, a block is stable when the realizations of $\mathbf{X}_t$ and $\mathbf{X}_{t+1}$ are identical. This happens if every component of $\tilde{\mathbf{X}}_t$ deviates from the corresponding one in $\mathbf{X}_t$ for an absolute amount lower than $1/2$, so that $[\tilde{\mathbf{X}}_t] = \mathbf{X}_t$. In other words, the probability that a block is stable in an image compressed t times with JPEG-100 is given by:

$$p_t \doteq \Pr\left(||\Delta\mathbf{X}_t||_\infty < \frac{1}{2}\right), \quad \Delta\mathbf{X}_t \doteq \tilde{\mathbf{X}}_t - \mathbf{X}_t. \tag{2}$$

By exploiting the linearity of the direct and inverse DCT, we obtain:

$$\Delta\mathbf{X}_t = \mathrm{IDCT2}(\Delta\mathbf{Y}_t), \quad \Delta\mathbf{Y}_t \doteq \tilde{\mathbf{Y}}_t - \mathbf{Y}_t, \tag{3}$$

thus the deviation of $\tilde{\mathbf{X}}_t$ from $\mathbf{X}_t$ is entirely due to the quantization operation in the DCT domain.

We can now exploit two observations:

1. $\Delta\mathbf{Y}_t$ is the quantization noise in the DCT domain, where the quantization step is always equal to 1. We assume that all the components of $\Delta\mathbf{Y}_t$ are independent with zero-mean and the same variance σ_t^2. It is justified by the fact that the distribution of the DCT coefficients resembles a Generalized Gaussian (even after high quality compression), and its variance is much larger than the quantization step.[1] Empirical evidence will be given in Section 4. A similar model is also assumed in [9] for the quantization noise in the spatial domain. Under this assumption, the covariance matrix of $\Delta\mathbf{Y}_t$ is diagonal with constant entries equal to σ_t^2:

$$\boldsymbol{\Sigma}_{\Delta\mathbf{Y}_t} = \sigma_t^2\mathbf{I}, \tag{4}$$

where $\mathbf{I}$ is the 64×64 identity matrix.

2. $\Delta\mathbf{X}_t$ is a linear transformation of $\Delta\mathbf{Y}_t$, which contains independent entries. In particular, due to the separability of the two-dimensional DCT and its inverse, we have that

$$\Delta\mathbf{X}_t = (\mathbf{D}\otimes\mathbf{D})^{-1}\Delta\mathbf{Y}_t, \tag{5}$$

where $\mathbf{D}$ is the 8×8 matrix of the 1-D DCT trasform and $\otimes$ indicates the matrix Kronecker product.
Thanks to the Central Limit Theorem (in Lindenberg's version), and under the assumption of independence in $\Delta\mathbf{Y}_t$, it follows that all the components of $\Delta\mathbf{X}_t$ are approximately zero-mean and jointly Gaussian. Moreover, the covariance matrix $\boldsymbol{\Sigma}_{\Delta\mathbf{X}_t}$ is

$$\begin{aligned}\boldsymbol{\Sigma}_{\Delta\mathbf{X}_t} &= (\mathbf{D}\otimes\mathbf{D})^{-1}\boldsymbol{\Sigma}_{\Delta\mathbf{Y}_t}((\mathbf{D}\otimes\mathbf{D})^{-1})^T \\ &= (\mathbf{D}^{-1}\otimes\mathbf{D}^{-1})\,\sigma_t^2\mathbf{I}\,(\mathbf{D}\otimes\mathbf{D}) = \sigma_t^2\mathbf{I},\end{aligned} \tag{6}$$

where we have exploited the orthogonality of the DCT transform ($\mathbf{D}\mathbf{D}^T = \mathbf{I}$) and the properties of the Kronecker product.

In other words, the components of the quantization noise vector $\Delta\mathbf{Y}_t$ are transformed through the inverse DCT into decorrelated and jointly normal (thus *independent*) components of $\Delta\mathbf{X}_t$ with the same mean 0 and variance σ_t^2. Then, we can write

$$\Delta\mathbf{X}_t[i] \sim \mathcal{N}(0, \sigma_t^2), \tag{7}$$

where $\mathcal{N}(\mu, \sigma^2)$ indicates a normal distribution with mean μ and variance σ^2. Consequently, we can derive p_t in (2) as follows

$$p_t = \prod_{i=1,\ldots,64}\Pr\left(|\Delta\mathbf{X}_t[i])| < \frac{1}{2}\right) = \left[\int_{-\frac{1}{2}}^{\frac{1}{2}}\mathrm{PDF}_{\mathcal{N}(0,\sigma_t^2)}(x)dx\right]^{64}, \tag{8}$$

where $\mathrm{PDF}_{\mathcal{N}(\mu,\sigma^2)}(x)$ is the probability density function of $\mathcal{N}(\mu, \sigma^2)$ evaluated in x.

It is worth pointing out that the value of p_t monotonically depends on the variance σ_t^2 of the quantization noise in the DCT domain: the higher is σ_t^2, the lower is p_t.

3.2. Block sample statistics

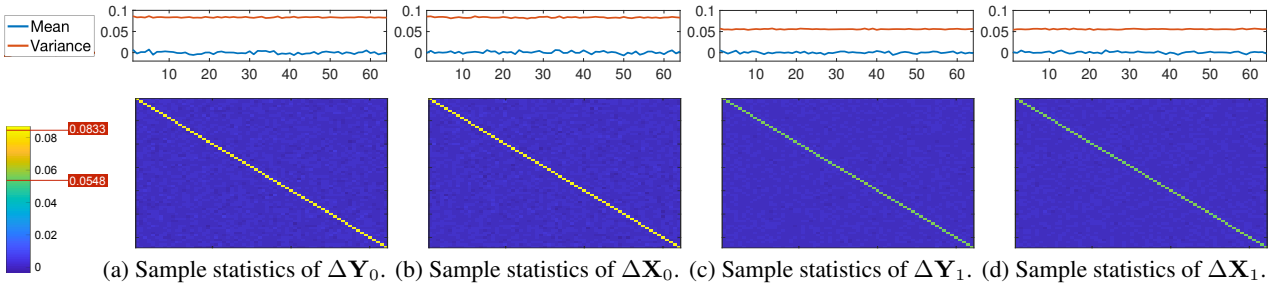
Practical forensic detectors based on [3, 5] consider a sample of N blocks of the image under investigation. The ratio of blocks that are stable after a JPEG-100 recompression is used as discriminative statistic in the forensic analysis. The number N depends on the size of the image region under investigation and on the pre-selection strategy. For instance, in [3] and [5] the authors suggest to discard flat blocks, i.e., blocks where all the pixels have the same value.

We denote as R_t the random variable representing this ratio, where t is always the number of previous JPEG compressions the image under analysis underwent. Let us indicate as $\Delta\mathbf{X}_t^{(n)}$ the random vector representing the n-th block in the sample, and define the binary random variables,

$$S_t^{(n)} = \begin{cases} 1 & \text{if } ||\Delta\mathbf{X}_t^{(n)}||_\infty < \frac{1}{2} \text{ (the } n\text{-th block is stable)} \\ 0 & \text{otherwise.}\end{cases} \tag{9}$$

According to Section 3.1, all of them have a Bernoulli distribution with parameter p_t, thus their expected value is p_t and their variance is $p_t(1-p_t)$. Moreover, we suppose that the $S_t^{(n)}$ are spatially

[1] Note that in case of quantization steps larger than 1, correlation in quantization noise might arise [8].



(a) Sample statistics of $\Delta\mathbf{Y}_0$. (b) Sample statistics of $\Delta\mathbf{X}_0$. (c) Sample statistics of $\Delta\mathbf{Y}_1$. (d) Sample statistics of $\Delta\mathbf{X}_1$.

Fig. 2. Sample statistics from 1 million 8×8 blocks $\mathbf{X}_0$ randomly generated and processed as in Fig. 1.

independent (i.e., for a fixed t). Then, R_t is the sample mean of these independent binary variables and, due to the Central Limit Theorem, it is approximately normal distributed around the true expected value p_t with variance inversely proportional to the number of blocks N:

$$R_t = \frac{\sum_{n=1}^{N} S_t^{(n)}}{N}, \quad \Rightarrow \quad R_t \sim \mathcal{N}_t, \tag{10}$$

where $\mathcal{N}_t \doteq \mathcal{N}\left(p_t, \frac{p_t(1-p_t)}{N}\right)$. Note that in practice R_t is a discrete random variable with a binomial distribution (scaled by factor $1/N$), which lends itself also to a (discrete) Poisson approximation. However, we use the normal approximation because dealing with continuous cdfs in more convenient, and it offers a better approximation with respect to the Poisson when p_t is not small (e.g., in the case of repeated compression).

4. APPLICATION IN A FORENSIC DECISION PROBLEM

The analysis developed in Section 3 enables a statistical characterization of the ratio of stable blocks in a grayscale image JPEG-100 compressed t times if the quantization noise variance is known. Thus, we need a model for σ_t^2 to characterize the distribution of R_t.

To instantiate our theory in a framework with testable hypotheses, we complement it with theoretical results on JPEG error analysis [1]. This work proposes upper bounds for σ_t^2 in the case of $t = 0$ (uncompressed images) and $t = 1$ (images JPEG-100 compressed once and then decompressed). While pointing the reader to [1] for details, we report the estimated bounds (from Eq. 7 of [1, p. 561])

$$\sigma_0^2 \leq 0.0833, \quad \sigma_1^2 \leq 0.0548, \tag{11}$$

which are obtained by exploiting the expression of the quantization noise variance for Gaussian and Laplace variables proposed in [10].

In Fig. 2, we show sample statistics of $\Delta\mathbf{Y}_0$, $\Delta\mathbf{X}_0$, $\Delta\mathbf{Y}_1$, $\Delta\mathbf{X}_1$ from 1 million randomly generated 8×8 blocks $\mathbf{X}_0$ with uniformly distributed integers between 0 and 255. For each column, the upper plots report for the different 64 components (along the horizontal axis) the sample mean and variance, which lie around zero and the two bounds in (11), respectively. The lower plots represent the sample covariance matrices (colors refer to the bar on the left), which clearly have a diagonal nature. These data confirmed that the assumptions in Section 3 are reasonable.

Then, we use a simplified model where an equality relationship holds in (11) to compute p_0 and p_1 as in (8), obtaining

$$p_0 = 0.003, \quad p_1 = 0.119. \tag{12}$$

Finally, in order to distinguish between never compressed and once JPEG-100 compressed images, we formulate the following hypothesis test:

$$H_0 : R \sim \mathcal{N}_0 \quad \text{vs} \quad H_1 : R \sim \mathcal{N}_1. \tag{13}$$

5. EXPERIMENTAL VALIDATION

In order to validate the theoretical models and explore their practical implications, we have conducted experimental tests on image patches of different size and content.

We have mentioned earlier that the assumption of negligible arithmetic errors of the DCT implementation needs scrutiny. In our tests we consider the widely used `libjpeg` library (version 6b). It supports three different DCT implementations: the *fast* method produces visible artifacts and is rarely used, the *float* method employs floating-point arithmetic and behaves more similarly to the textbook DCT (matrix multiplication). The *slow* method relies on integer arithmetic. All three implementations use a divide-and-conquer algorithm that reuses intermediate results and may amplify small initial errors. We exclude the first method from our analysis and focus on the remaining two, keeping in mind that the *float* method likely follows more closely the theoretical model but the *slow* method is set as default in `libjpeg`, thus being more widely used.

We took 1488 never-compressed images from the Dresden dataset [11] and 500 images from the RAISE-1K dataset [12]. We have created a subset of images, denoted as $\mathcal{I}_0$, fulfilling the null hypothesis by converting the 1988 color images to grayscale and saving them in uncompressed format. For the alternative hypothesis, the images have been compressed with JPEG-100, using *float* and *slow* DCT implementations, giving the sets $\mathcal{I}_{1,F}$ and $\mathcal{I}_{1,S}$, respectively.

We classified squared patches of different size (256×256, 128×128, 64×64, 32×32) coming from the test set images. For each patch size, 19 880 patches (10 per image) have been randomly extracted, processed, and classified as follows:

1. Flat blocks are identified and only the remaining N non-flat blocks are considered, as suggested in [3, 5].
2. The patch is JPEG-100 recompressed by using the specified DCT implementation.
3. Each block among the N non-flat ones is compared to its recompressed counterpart: if the n-the block is stable, then $s^{(n)}$ (the realization of $S^{(n)}$) is 1, otherwise it is 0.
4. The ratio $r = \frac{\sum_{n=1}^{N} s^{(n)}}{N}$ is computed and the classification is performed with a Maximum Likelihood rule:[2] the $\hat{i}$-th hypothesis is accepted iff

$$\hat{i} = \arg \max_{i \in \{0,1\}} \text{PDF}_{\mathcal{N}_i}(r).$$

[2]It is to be noted that applying rule 4 is equivalent to performing a Neyman-Pearson test and fixing the threshold of the likelihood ratio test to 1.

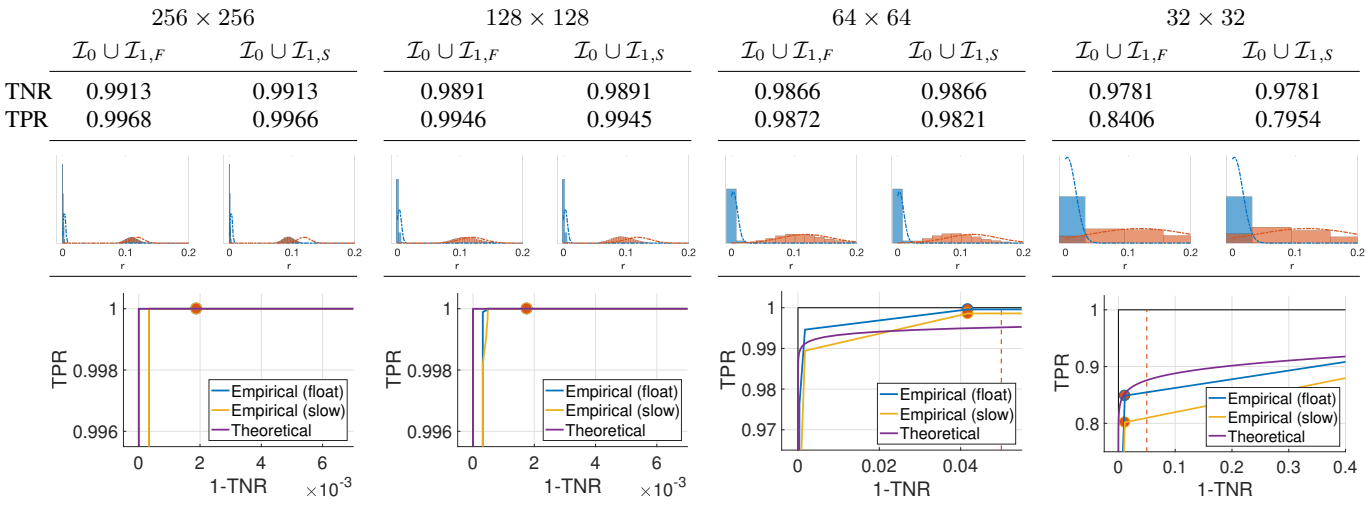


Fig. 3. Classification results for different patch size.

Figure 3 reports the results from the test set. Subtables refer to different patch sizes. In each subtable, we report in the results obtained with the ML rule and we differentiate according to the DCT implementation considered for images verifying H_1, by arranging the two cases column-wise. True Negative Rates (TNR) and True Positive Rates (TPR) are reported in each case, where the former is constant across columns as $\mathcal{I}_0$ is the same in both cases. Below each subtable, we also plot the histograms of the ratios r obtained from patches in $\mathcal{I}_0$ (blue bins), $\mathcal{I}_{1,F}$ (orange bins in the left plot) and $\mathcal{I}_{1,S}$ (orange bins in the right plot), together with the theoretical pdfs. Here, only patches for which all the blocks are non-flat have been considered, so that N is the same for every value of r at a fixed patch size, and bins are centered at n/N, $n = 0, \dots, N$. Additionally, for these patches we also propose a comparison of ROC curves: we plot the theoretical one (computed starting from the theoretical cdfs) and the empirical ones obtained from the data, by differentiating again the two DCT methods. In red, we report the point on the ROC curves corresponding to the threshold with significance level 0.05, that is theoretically determined from $\mathcal{N}_0$. The dashed red line indicates this false positive rate equal to 5% (if visible on the scale).

As expected, the more blocks are involved in the computation of r, the better the data are separated. This behavior is well modeled by the theoretical distributions and leads to higher accuracies when the patch size increases. It is however noticeable that the actual distribution of the ratio is often a bit shifted to the left: it is essentially negligible when the patch size is small and/or the *float* DCT is employed, while it is quite evident for the *slow* DCT in bigger patches where the distribution has a lower dispersion around the mean. We conjecture the stronger rounding and approximations due to integer arithmetic arising in the *slow* implementation are no longer negligible. A more significant deviation from the simplified model of the quantization noise variance assumed is introduced, thus calling for a more accurate model for σ_t^2 (and, consequently, for p_t).

Nevertheless, the detection results by means of the ML rule are robust across the two different DCT implementation, with minimal impact on the TPR. Accuracies are above 98.43% in all cases, except for 32×32 patches where a TPR drop is encountered, as explained by the more pronounced overlap of the theoretical pdfs.

Moreover, by looking at the ROC curves, we observe that the theoretical thresholds at 0.05 significance level always yield a false positive rate lower than 5% and the highest possible TPR in the area delimited by the red dashed line. In other words, the theoretical cdfs provide the best threshold in terms of TPR when allowing a false positive rate of 5%. Calibrating the threshold from our empirical data does not lead to any benefit. Tests performed by adopting the binomial model (not reported here for the sake of brevity) do not show significant differences, although p_0 is close to 0. This confirms that our theory is practical.

6. CONCLUSION

We have proposed a theoretical approach to JPEG block convergence for grayscale images and maximum quality compression.

The approximate distribution of the stable block ratio at different compression stages has been derived as a function of the quantization noise variance in the DCT domain. Tests on images with different content and size confirm the validity of the model and its capability to enable accurate hypothesis tests. In fact, the discrimination of never compressed and decompressed images after JPEG-100 compression yields an accuracy ranging from 98.4% to 99.6% by means of a calibration-free Maximum Likelihood rule. We also found that calibration from empirical data does not improve performance when fixing a false alarm rate upper bound.

While the presented results are encouraging, they represent only a step towards the full comprehension of JPEG block convergence. Several directions should be explored in future work. This can include the extension to color images and to other quality factors. As pointed out in Section 5, the quantization noise variance model could be adapted in order to better represent the effect of the *slow* DCT implementation, used as default in popular libraries. This could also enable a theoretically founded identification of the DCT method, as done empirically in [3]. Finally, a systematic way to compute the quantization noise variance after several compressions would allow us to model the full convergence path of a generic block beyond the first two compression cycles.

Acknowledgment. This research was funded by Deutsche Forschungsgemeinschaft (DFG) under grant "Informationstheoretische Schranken digitaler Bildforensik" and by Archimedes Privatstiftung, Innsbruck.

7. REFERENCES

[1] 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.

[2] J. Yang, J. Xie, G. Zhu, S. Kwong, and Y.Q. Shi, "An effective method for detecting double JPEG compression with the same quantization matrix," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 11, 2014.

[3] Shi-Yue Lai and Rainer Böhme, "Block convergence in repeated transform coding: JPEG-100 forensics, carbon dating, and tamper detection," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2013, pp. 3028–3032.

[4] K. Wang A. T. P. Ho and F. Cayre, "An effective histogram-based approach to JPEG-100," in *IEEE International Conference on Image Processing Theory Tools and Applications (IPTA)*, 2016.

[5] Matthias Carnein, Pascal Schöttle, and Rainer Böhme, "Forensics of high-quality JPEG images with color subsampling," in *IEEE Workshop on Informations Forensics and Security (WIFS)*, 2015.

[6] Cecilia Pasquini, Pascal Schöttle, Rainer Böhme, Giulia Boato, and Fernando Pèrez-Gonzàlez, "Forensics of high quality and nearly identical JPEG image recompression," in *ACM Information Hiding and Multimedia Security Workshop*, Vigo, Galicia, Spain, 2016, pp. 11–21.

[7] B. Li, T.-T. Ng, X. Li, S. Tan, and J. Huang, "Statistical model of JPEG noises and its application in quantization step estimation," *IEEE Transactions on Image Processing*, vol. 24, no. 5, pp. 1471–1484, 2015.

[8] M.A. Robertson and R. L. Stevenson, "DCT quantization noise in compressed images," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 1, pp. 27–38, 2005.

[9] D. Vàzquez-Padìn, F. Pèrez-Gonzàlez, and P. Comesaña-Alfaro, "A random matrix approach to the forensic analysis of upscaled images," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 9, pp. 2115–2130, 2017.

[10] A. B. Sripad and D. Snyder, "A necessary and sufficient condition for quantization errors to be uniform and white," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 25, no. 5, pp. 442–448, 1977.

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

[12] 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.