Forensics of high quality and nearly identical JPEG image recompression

C. Pasquini
University of Trento, Italy
cecilia.pasquini@unitn.it

P. Schöttle
Universität Innsbruck, Austria
pascal.schoettle@uibk.ac.at

R. Böhme
Universität Innsbruck
rainer.boehme@uibk.ac.at

G. Boato
University of Trento, Italy
boato@disi.unitn.it

F. Pèrez-Gonzàlez
University of Vigo, Spain
fperez@gts.uvigo.es

ABSTRACT

We address the known problem of detecting a previous compression in JPEG images, focusing on the challenging case of high and very high quality factors (≥ 90) as well as repeated compression with identical or nearly identical quality factors. We first revisit the approaches based on Benford–Fourier analysis in the DCT domain and block convergence analysis in the spatial domain. Both were originally conceived for specific scenarios. Leveraging decision tree theory, we design a combined approach complementing the discriminatory capabilities. We obtain a set of novel detectors targeted to high quality grayscale JPEG images.

Keywords

JPEG forensics; High quality images; Double compression.

1. INTRODUCTION

Reconstructing of the compression history of JPEG images is relevant in multimedia forensics. Traces of prior compression may indicate multiple processing steps, potentially including forgeries. In recent years the problem of detecting traces of multiple compressions has been extensively studied. This has resulted in a number of forensic approaches targeted to detect double or multiple compression.

The known approaches face some important limitations. First, multiple compression detectors are generally evaluated by considering quite strong quantization. Typically, the tested JPEG quality factors (QF) are lower than 90. However, high and very high quality JPEG compression is relevant as memory and bandwidth are cheap. One may also speculate that counterfeiters store intermediate versions at high quality to avoid visible artifacts and detectable traces. Second, cases where the second compression has a lower quality than the prior compression remain hard to detect with the known methods. Third, the majority of techniques is based on DCT coefficients. These techniques typically fail in detecting previous compressions performed with the same quantization matrix. While some progress has been made in this direction [6, 16, 7, 3], a combination of these techniques with the ones based on DCT coefficients remains unexplored.

In this paper, we close this gap. We focus on the specific scenario where both the primary and secondary quality factors are larger or equal to 90. We aim at detecting a previous compression also in those cases where the quantization matrix used in the secondary compression is identical or nearly identical to the primary one. In order to achieve this goal, we consider two different and recently proposed approaches: the one based on block convergence in the spatial domain [7, 3] and the one based on Benford–Fourier analysis [10]. After assessing the strengths and limitations on each of them in the considered scenario, we proposed to combine them. Decision tree theory offers a general framework to do this systematically. This results in a set of double compression detectors for high quality JPEG images, parameterized by the observable quality factor of the image under investigation. We show that the combined classification tests lead to considerable benefits in terms of classification accuracy compared to the state of the art.

This work introduces several contributions: we improve both state of the art detectors individually and adapt them to high and very high JPEG qualities. For Benford–Fourier we propose to adaptively choose the DCT frequencies used in the test in order to accommodate for the fact that quantization matrices for nearly identical quality factors do not differ in every coefficient. We simplify the depth of the block convergence analysis to the relevant number of iterations. We build and evaluate a joint detector that outperforms each of the improved methods individually.

The paper is structured as follows: Section 2 discusses the decision problems studied in the literature and proposes terminology to precisely define the tests in this paper. Section 3 describes the two forensic techniques we consider for our joint detector. In Section 4, we measure the performance of both techniques in the considered scenario and improve them. Section 5 presents the combination approach and reports the performance of the combined detectors. Finally, Section 6 concludes with an outlook on future work.

ACM IH&MMSec 2016 Vigo, Spain
© 2016 ACM. ISBN 978-1-4503-4290-2/16/06...$15.00
DOI: http://dx.doi.org/10.1145/2909827.2930787

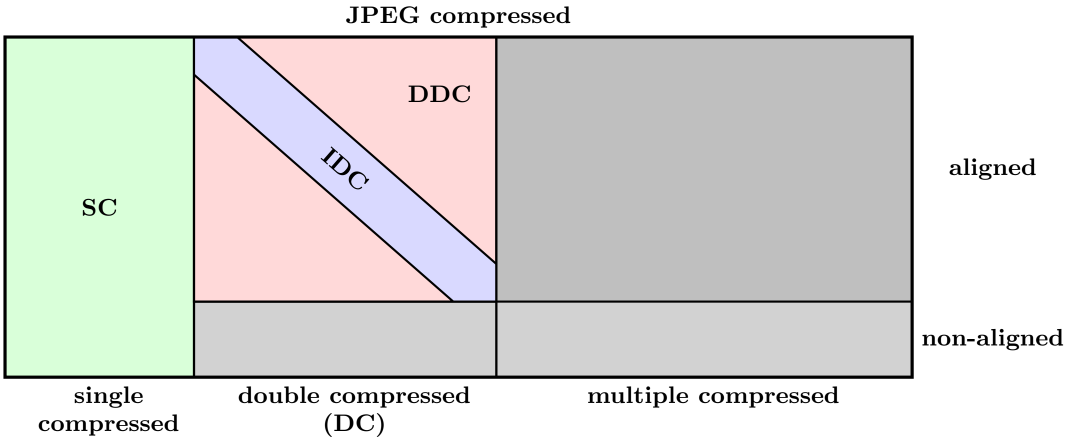


Figure 1: Representation of the multiple JPEG compression detection problem

2. PROBLEM STATEMENT

Figure 1 depicts the possible status of a JPEG image. Either the image was compressed only once and belongs to the strict subset of single compressed JPEG images (**SC**), or it was compressed multiple times. In the latter case, (some of) the compressions could be non-aligned (light gray subset) or all of them can be aligned (dark gray subset). In this paper we focus on aligned compressions, thus we do not depict the non-aligned set in full detail. Both subsets, aligned and non-aligned, can be further distinguished into the subset of images that were compressed exactly twice (**DC**) and the subset of images with more than two compressions. Again, we concentrate on the relevant case for this paper. Aligned and double compressed JPEG images can be further divided into those that were compressed twice with the same quantization matrix (**IDC**) and those that were the quantization matrices differ between the two compressions (**DDC**). It is known that different quantization tables are generated as function of the quality factor (QF), a natural number between 1 and 100. The lower the quality factor, the higher are the values in the quantization table. Then, further distinctions are possible. For instance, **DDC** could be splitted into the set where the QF of the first compression was strictly higher or strictly lower than the QF of the second compression.

Ideally, a general framework for detecting multiple JPEG compressions should differentiate between all of the depicted subsets, although most of the approaches in the literature restrict themselves to the distinction of some of the subsets. One of the most studied problem is certainly the distinction between **SC** and **DC** images in the aligned case, which is addressed in [13, 1, 4, 8], while some approaches target also a multiple recompression [9, 10]. However, most of them fail when dealing with images in **IDC**, thus specific approaches have been designed for this case, targeting also the detection of multiple recompression with the same quantization matrix [7, 3, 6, 16].

Given the diversity of statistical traces in **DDC** and **IDC** images, in the proposed approach we reformulate the **SC** vs **DC** discrimination problem as the distinction of the three different sets **SC**, **DDC** and **IDC**. Moreover, we focus on the case of QF≥ 90 both for the primary and secondary compression and indicate such scenario as *HQ-DC* (high quality double compression) for the sake of clarity.

3. CURRENT APPROACHES

In this section, we will briefly recall the main rationale and procedures of the two state-of-the-art approaches that are leveraged in this work, stressing the different forensic scenarios for which they were originally designed.

3.1 Benford–Fourier analysis

The Benford–Fourier (BF) coefficients have been introduced in [12] and they proved to be particularly suitable for the analysis of DCT coefficients, thus providing an effective tool for different tasks in image forensics. In particular, given a continuous random variable Z with pdf $f_Z(z)$, its *Benford–Fourier coefficient* at $\omega \in \mathbb{R}$ is defined as

$$a_\omega = \int_{-\infty}^{+\infty} f_Z(z) e^{-j\omega \log_{10} z} dz.$$

In [11] and [10], BF coefficients have been used to study the distribution of 8×8 DCT coefficients for the detection of previous compressions in uncompressed format and JPEG format images, respectively. In particular, given the analyzed image the BF coefficients are estimated by the formula

$$\hat{a}_\omega = \frac{\sum_{m=1}^{M} e^{-j\omega \log_{10} z^m}}{M} \qquad m = 1, \ldots, M \tag{1}$$

where z^m is the realization of the random variable representing the absolute value of DCT coefficients at a certain frequency in the m-th block of the image, M is the total number of 8×8 blocks.

In particular, in [10] such estimates are computed for a number of DCT frequencies in a subset $F \subset \{1, \ldots, 64\}$. Specifically, there is a BF coefficient at certain ω^f for each $f \in F$, thus obtaining a vector

$$\hat{\mathbf{a}} = [\hat{a}^1_{\omega^1}, \ldots, \hat{a}^F_{\omega^F}].$$

After an estimation of their pdf in the different quantization chains, the elements in $\hat{\mathbf{a}}$ are combined to compute the likelihood function $\mathcal{L}(H|\hat{\mathbf{a}})$ for a certain hypothesis of compression history H. The null hypothesis H_N is given by the fact that the last JPEG compression (whose parameters are known from the image under investigation) is the only one that occurred in the digital history of the image, while the alternative hypotheses depend on the forensic scenario considered (in [10], single and triple compression are analized). For instance, in a double compression detection framework each alternative hypothesis H_A is represented by the fact that the image was previously compressed with a primary quality factor among a predefined set. For each H_A, the logarithmic likelihood ratio (LLR) is computed as follows

$$\mathrm{LLR} = -2 \cdot \ln \left[\frac{\mathcal{L}(H_N|\hat{\mathbf{a}})}{\mathcal{L}(H_A|\hat{\mathbf{a}})} \right], \tag{2}$$

and the maximum value of the LLR over all the alternative hypotheses, indicated as LLR_m, is considered. If $\mathrm{LLR}_m > 0$ then the null hypothesis is rejected, otherwise it is accepted.

In [10], the method has been tested on single and double compressed images created by combining the quality factors in $\{50, 60, 70, 80, 90\}$, i.e., the difference between the primary and secondary quality factor is in any case at least 10. It provided good accuracies, which are generally maintained in the cases of stronger secondary compression but dramatically drops in the case of identical recompression.

3.2 Block convergence

In [7], the authors propose a technique to identify the number of JPEG compressions with quality factor 100 in grayscale images. In this case the quantization table is composed only of the value 1 and we will indicate such setting as JPEG-100. It is observed that for JPEG-100 the 8 × 8 blocks are transformed in the DCT domain, rounded to the nearest integer and transformed back to the pixel domain, where they are again rounded to the nearest positive integer and truncated to the value range. The authors show that for some of the blocks none of the pixel values change during a JPEG-100 compression. They call these blocks stable and conjecture that after repeated JPEG-100 compression, all blocks of a grayscale image will converge, i.e., become stable. Furthermore, the percentage of blocks that becomes stable after a certain number of recompressions is largely independent of the image content. In particular, given a subject JPEG-100 image, the authors propose to discard the flat blocks (i.e., the ones which contain a single value and are stable from the beginning) and, among the remaining ones, count the ones that become stable after each JPEG-100 recompression. Hence, the ratio of stable blocks (for different numbers of JPEG compressions) is computed by:

$$r = \frac{b_{\text{stable}} - b_{\text{flat}}}{b_{\text{total}} - b_{\text{flat}}}, \tag{3}$$

where b_{total} is the total number of blocks in the image, b_{flat} is the number of flat ones and b_{stable} is the number of stable ones. The value of r is then used to identify the number of previous JPEG-100 compressions.

This approach has been extended in [3] to color images, for which all the three color channels need to be analyzed and the block convergence path is studied with respect to a number of additional aspects, such as the kind of color space conversion and the subsampling/upsampling methods. Moreover, in this work the authors propose to fit a theoretical distribution (specifically, the beta distribution) to the ratios of stable blocks observed after different numbers of recompressions.

The technique has been tested on images that were recompressed multiple times with the very same quality factor and provided accurate results in case of very high quality images (QF$\in \{100, 99\}$). It also has been noted that the accuracy decreases together with the quality factor.

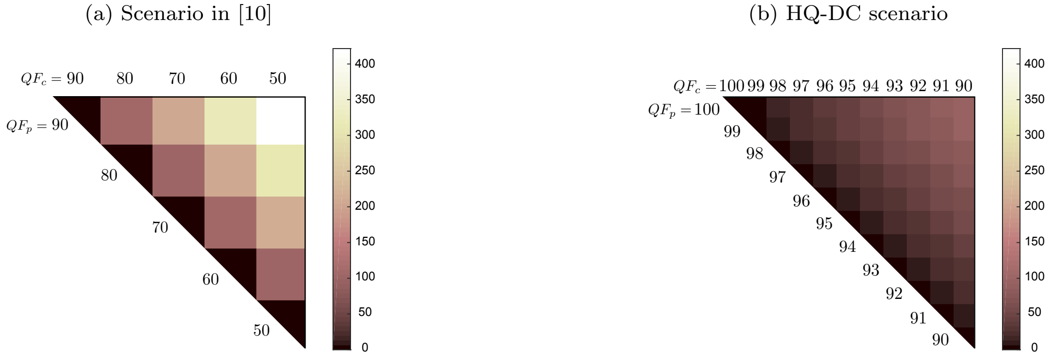
4. INDIVIDUAL APPROACHES

As already mentioned in Section 1, we focus on the detection of high quality double compression in JPEG images, i.e. with quality factors higher or equal to 90. In this section we first describe our set-up and then show the performance of the techniques mentioned in the last section in our scenario. Furthermore, we improve both methods.

4.1 HQ-DC scenario

In our forensic scenario, each grayscale test image has a *current quality factor* $\mathrm{QF}_c \in \{90, 91, \ldots, 100\}$ which is known from the given JPEG file and a *previous quality factor* $\mathrm{QF}_p \in \{90, 91, \ldots, 100, \mathrm{NC}\}$, where we will use the notation QF_p =NC if the image is single compressed and has no primary quality factor. We will represent a JPEG compression history as square brackets containing the ordered sequence of quality factors applied. Thus, in our tests the compression history of the image under investigation can be either [NC, QF_c] (only the last JPEG compression occurred) or [QF, QF_c] (a previous compression occurred), where QF is searched within set $\mathcal{QF}_{range} = \{90, 91, \ldots, 100\}$.

This kind of setting implies a limited difference between the quantization tables used in the primary and secondary


(a) Scenario in [10]

(b) HQ-DC scenario

Figure 2: Euclidean norm of the quantization table difference in the different double compression chains

	NC	100	99	98	97	96	95	94	93	92	91	90
100	🟩	🟦	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
99	🟩	🟥	🟦	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
98	🟩	🟥	🟥	🟦	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
97	🟩	🟥	🟥	🟥	🟦	🟥	🟥	🟥	🟥	🟥	🟥	🟥
96	🟩	🟥	🟥	🟥	🟥	🟦	🟥	🟥	🟥	🟥	🟥	🟥
95	🟩	🟥	🟥	🟥	🟥	🟥	🟦	🟥	🟥	🟥	🟥	🟥
94	🟩	🟥	🟥	🟥	🟥	🟥	🟥	🟦	🟥	🟥	🟥	🟥
93	🟩	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟦	🟥	🟥	🟥
92	🟩	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟦	🟥	🟥
91	🟩	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟦	🟥
90	🟩	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟦

Table 1: Example of result representation. Each cell refers to a compression chain specified by the quality factors at the corresponding row and column. Current quality factors QF$_c$ are reported row-wise and primary ones are reported column-wise. Colors match the ones used in Figure 1.

JPEG compression, as well as generally small primary quantization steps which makes the detection of a previous quantization more difficult. Figure 2 shows the Euclidean norm of the differences between the quantization tables (luminance channel) of the quality factors considered for the experimental setting in [10], compared with the ones of the HQ-DC scenario. We refer to the standard quantization tables used by the `libjpeg` library released by the IJG (Independent JPEG Group), as they are often used in common software.

With respect to the scenario described in Section 2, our goal is to design a test that analyzes a JPEG image with QF$_c \geq 90$ either belonging to **SC**, **DDC**, or **IDC** and assign it to the correct set[1]. Table 1 serves an example of the result representation for our experimental setting: green cells refer to compression chains of the form [NC,QF$_c$] (class **SC**), blue cells to compression chains of the form [QF$_c$,QF$_c$] (class **IDC**), and red cells to compression chains of the form [QF,QF$_c$], QF≠QF$_c$ (class **DDC**). In the following, the numbers in each cell will indicate the classification accuracy (i.e., the percentage of images that are assigned to the correct class) with respect to the specific test considered.

For our experiments, we consider two datasets composed of images in TIFF format and widely used in the literature: the UCID database [15] (1338 images, 384×512) and a subset of the DRESDEN database [5] (1488 images, 3872×2592 and 3008×2000). For each image we create single and double compressed versions by combining all the quality factors in $\mathcal{QF}_{range}$. Thus, each image is processed according to 132 different compression chains, 11 with single compression and 121 with double compression. We limit our analysis to grayscale images and consequently apply the approaches to the luminance channel only. We employ the state-of-the-art libraries `libtiff 3.6.1` and `libjpeg 8d` to read TIFF and write grayscale JPEG images, respectively.

In this section, for the sake of brevity, we report the results for the UCID database only, as those for the DRESDEN database are very similar. We reproduced the experimental setting used in [10], where a set of 600 UCID images was used for estimating the prediction error parameters (which are employed also for other datasets). Such images are then excluded from the Benford–Fourier analysis, while the remaining ones are used for testing.

4.2 Benford–Fourier analysis

We first consider the method proposed in [10], that we indicate as `BF_baseline`. Here, a predefined set of 9 DCT frequencies (specifically $F = \{4, 6, 11, 13, 15, 22, 24, 26, 28\}$ in zigzag order) is used to compute (2). Moreover, the potential primary quality factor is searched within the whole set $\mathcal{QF}_{range} = \{90, 91, \ldots, 100\}$, thus each image is classified either as belonging to **SC** (if the maximum value of LLR is below 0), **DDC** (if the estimated compression chain [QF,QF$_c$] is such that QF≠QF$_c$) or **IDC** (if the estimated

	NC	100	99	98	97	96	95	94	93	92	91	90
100	0.05	0.98	0.00	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
99	0.01	0.99	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
98	0.08	0.04	0.03	0.91	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
97	0.21	0.07	0.07	0.65	0.81	1.00	1.00	1.00	1.00	1.00	1.00	1.00
96	0.20	0.03	0.02	0.26	0.90	0.81	0.99	1.00	1.00	1.00	1.00	1.00
95	0.24	0.04	0.08	0.21	0.69	0.71	0.76	0.95	1.00	1.00	1.00	1.00
94	0.27	0.02	0.03	0.41	0.62	0.97	0.78	0.74	0.91	1.00	1.00	1.00
93	0.43	0.06	0.07	0.25	0.71	0.88	0.98	0.83	0.57	0.99	0.99	1.00
92	0.57	0.01	0.04	0.03	0.49	0.17	0.90	0.94	0.77	0.41	0.81	0.96
91	0.54	0.12	0.14	0.75	0.80	0.89	0.98	1.00	0.95	0.76	0.37	0.86
90	0.69	0.05	0.06	0.11	0.26	0.72	0.78	1.00	1.00	0.98	0.71	0.27

Table 2: Accuracy of the `BF_baseline` test.

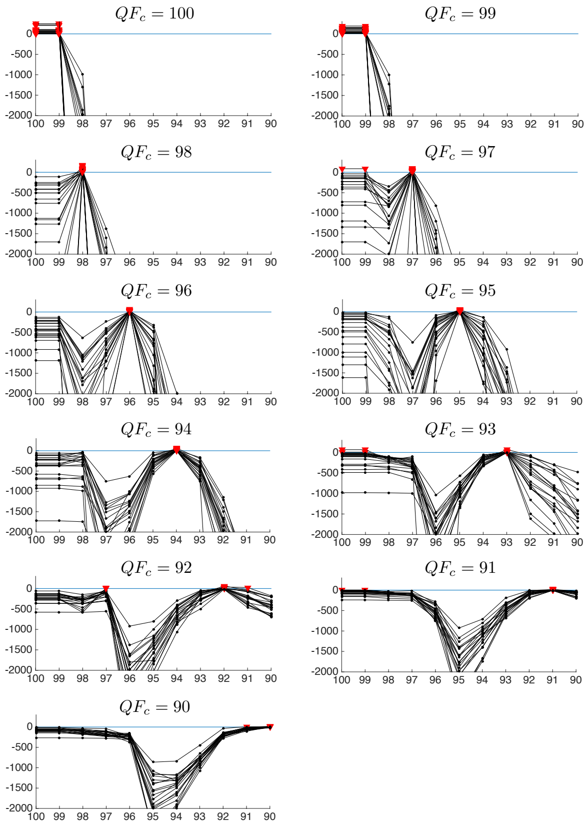


Figure 3: Values of the LLRs among different QF$_c$ for 20 randomly selected UCID images. In each plot the horizontal axis contains the quality factors in $\mathcal{QF}_{range}$. The vertical axis represents the value of LLR, that we report in the interval $[-2000, 200]$ in order to compare its behavior across the different QF$_c$. Each black line corresponds to an image and each value of the LLR that lies above 0 is marked in red.

[1] With a slight abuse of notation, we will indicate as **SC**, **DDC** and **IDC** the sets introduced in Section 2, although they only refer to the case of high quality JPEG compression (QF≥ 90).

compression chain is such that QF $=$ QF$_c$). However, the performance of `BF_baseline` degrades when moving from the scenario in Fig. 2a to the HQ-DC one, as reported in Table 2.

We notice that, although the accuracy for **DDC** is generally high when QF$_p$ <QF$_c$ (the upper triangle of the table), we have a substantial misclassification for **SC**. By exploring the results more closely, we obtain that the Benford–Fourier analysis for images in **SC** generally leads to values of LLR higher than 0 when the alternative hypothesis is given by [QF$_c$,QF$_c$]. This can be observed in Fig. 3, where we report the values of the different LLRs yielded by the quality factors in $\mathcal{QF}_{range}$ for single compressed images.

Moreover, we have high values of the LLR when the primary quality factor tested is high or close to QF$_c$, even for lower current quality factors. The former phenomenon is due to the small steps used in the primary quantization. The latter one is caused by the fact that the quantization tables of QF$_c$ and the one tested might share the very same quantization steps for all or some of the DCT frequencies used in the computation of the LLR, thus decreasing the distinguishability of the two hypotheses. For instance, the quantization table of 99 is equal to 1 up to the 37-th frequency and it fully coincides with the one of 100 at the DCT frequencies used for the computation of the LLR. Thus, when analyzing a JPEG image with QF$_c = 100$, the hypothesis $[99, 100]$ will yield the very same LLR as $[100, 100]$. This likely causes misclassification.

In order to cope with this issue, we propose to adaptively select the set F according to the binary hypothesis test, i.e., choosing the first 9 DCT frequencies in zigzag order among the ones that actually have different primary quantization steps. In the case of $[NC, 100]$ vs $[99, 100]$, the algorithm will choose the frequencies $\{37, 38, 41, 45, 46, 47, 48, 49\}$, where quantization steps for 99 are equal to 2. This implicitly forces to exclude the hypothesis [QF$_c$,QF$_c$] from the pool of alternative ones (as no suitable DCT frequencies would be identified) and to set $\mathcal{QF}_{range} = \{90, 91, \ldots, 100\} \setminus QF_c$. By this, we reduce the misclassification for single compressed images while being aware that the possibility of identical double compression needs to be assessed. In other words, we can design a new BF test, that we will denote as `BF_adaptive`, that has two possible outcomes: the image belongs either to the union **SC** $\cup$ **IDC** (LLR is below 0 for every hypothesis) or to **DDC** (at least one hypothesis has a LLR higher than 0).

We report in Fig. 4 an example of the different values of LLR obtained with the two different tests, where we can notice the benefit of the frequency selection. Moreover, the accuracy results of this approach are reported in Table 3. They show that the accuracy for **DDC** is unaltered with respect to the baseline approach. Misclassification for **SC** is now reduced, although it is no longer distinguished from **IDC** (for this reason the accuracy on **IDC** is also very high). On the other hand, the lower triangle of the table (especially when QF$_p$ > 95) remains an issue.

Thus, we can conclude that the Benford–Fourier analysis with adaptive selection of the DCT frequencies is suitable to detect non-identical double compression and is particularly accurate when QF$_p$ <QF$_c$ or QF$_p \leq 95$. This suggests that other techniques can be used to extend the analysis to the detection of identical recompression.

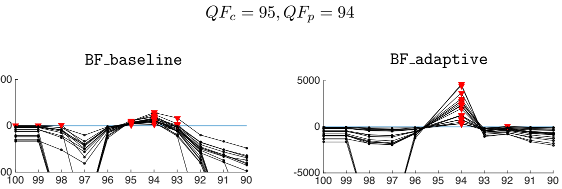


Figure 4: Effect of the adaptive DCT selection on the computation of LLR in case of a double compressed image with QF$_p = 94$ (for the `BF_adaptive` test the LLR for $[QF_c, QF_c]$ is not available).

	NC	**100**	**99**	**98**	**97**	**96**	**95**	**94**	**93**	**92**	**91**	**90**
100	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
99	0.95	0.13	0.96	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
98	0.94	0.14	0.18	0.94	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
97	0.91	0.17	0.22	0.79	0.91	1.00	1.00	1.00	1.00	1.00	1.00	1.00
96	0.94	0.11	0.12	0.47	0.88	0.95	1.00	1.00	1.00	1.00	1.00	1.00
95	0.92	0.13	0.16	0.36	0.85	0.95	0.93	0.99	1.00	1.00	1.00	1.00
94	0.95	0.15	0.18	0.20	0.20	0.98	0.99	0.94	0.99	1.00	1.00	1.00
93	0.93	0.16	0.19	0.35	0.32	0.96	0.99	0.99	0.93	0.99	1.00	1.00
92	0.94	0.15	0.20	0.24	0.23	0.53	0.96	0.99	0.99	0.94	0.99	1.00
91	0.91	0.16	0.18	0.69	0.79	0.83	0.99	1.00	1.00	0.99	0.90	1.00
90	0.93	0.14	0.19	0.20	0.38	0.44	0.94	1.00	1.00	1.00	0.99	0.93

Table 3: Accuracy of the `BF_adaptive` test

4.3 Block convergence

We compute for each UCID image the ratio r as in Equation (3), by recompressing the image with the current quality factor QF$_c$. Then, we used the 600 UCID images discarded in the Benford–Fourier analysis for fitting theoretical models, while the remaining ones are used for evaluating the different tests designed (i.e., the results in the tables refer to the very same images for both methods).

As a first approach, we adopt a maximum likelihood test as proposed in [3], which searches among the pool of potential primary quality factors $\mathcal{QF}_{range} = \{90, 91, \ldots, 100\}$ and is based on a theoretical approximation of the empirical data distribution. In particular, we fit a beta distribution for each of the 132 different compression chains and design a first discrimination test, that we will indicate as `BC_ML`, consisting of the following steps:

- Given an image with a certain QF$_c$, the value of r is computed.
- We consider all the beta distribution pdfs $p_{[\mathrm{QF},\mathrm{QF}_c]}(\cdot)$ that were previously estimated from every compression chain [QF, QF$_c$], where QF varies in $\mathcal{QF}_{range} \cup \{NC\}$.
- We evaluate each pdf for r and pick the one that yields the maximum value: if it corresponds to a QF $\in \mathcal{QF}_{range}$ then the image is classified as double compressed, while if it corresponds to the case of NC it is classified as single compressed.

It has to be pointed out that the `BC_ML` is in principle able to distinguish between the three different sets **SC**, **DDC** and **IDC**, as is the `BF_baseline`. However, it also has problems in accurately classifying **SC**, as shown in Table 4.

On the other hand the double compressed images are usually correctly identified, in both the **DDC** and **IDC** set (for QF$_c \geq 94$). We can identify the reason of the misclassification for **SC** by looking at the estimated beta pdfs reported in Fig. 5. Notice that they are strongly overlapping

	NC	**100**	**99**	**98**	**97**	**96**	**95**	**94**	**93**	**92**	**91**	**90**
100	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.99	0.99
99	0.78	0.53	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.98	0.98	0.98
98	0.75	0.27	0.28	0.99	0.76	0.98	0.96	0.97	0.98	0.98	0.99	0.98
97	0.26	0.75	0.76	0.81	0.99	0.95	0.96	0.99	0.98	0.98	0.99	0.99
96	0.38	0.64	0.63	0.69	0.76	0.98	0.91	0.88	0.88	0.97	0.95	0.94
95	0.01	0.99	0.99	0.99	0.99	0.99	0.98	1.00	0.99	0.99	0.99	1.00
94	0.12	0.90	0.89	0.89	0.89	0.88	0.91	0.97	0.95	0.93	0.92	0.89
93	0.10	0.90	0.90	0.90	0.90	0.89	0.90	0.92	0.89	0.96	0.94	0.94
92	0.17	0.82	0.81	0.82	0.82	0.82	0.83	0.83	0.83	0.33	0.89	0.88
91	0.19	0.80	0.80	0.80	0.80	0.80	0.79	0.81	0.80	0.81	0.16	0.88
90	0.07	0.88	0.89	0.89	0.91	0.90	0.90	0.91	0.89	0.89	0.91	0.18

Table 4: Accuracy of the `BC_ML` test

	NC	**100**	**99**	**98**	**97**	**96**	**95**	**94**	**93**	**92**	**91**	**90**
100	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
99	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.99	0.98	0.98	0.98
98	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99
97	1.00	1.00	1.00	1.00	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00
96	1.00	1.00	1.00	1.00	1.00	0.97	1.00	1.00	1.00	1.00	1.00	1.00
95	1.00	1.00	1.00	1.00	1.00	1.00	0.96	1.00	1.00	1.00	1.00	1.00
94	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.96	1.00	1.00	1.00	1.00
93	0.75	0.72	0.72	0.71	0.69	0.77	0.70	0.67	0.96	0.51	0.59	0.61
92	0.03	0.03	0.03	0.02	0.03	0.02	0.03	0.02	0.01	0.99	0.01	0.01
91	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.99	0.00
90	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.99

Table 5: Accuracy of the `BC_threshold` test with $t = t_1$

	NC	**100**	**99**	**98**	**97**	**96**	**95**	**94**	**93**	**92**	**91**	**90**
100	0.99	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
99	0.98	0.96	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
98	0.98	0.98	0.98	1.00	0.72	0.16	0.17	0.16	0.11	0.10	0.05	0.09
97	0.99	0.99	0.99	0.97	1.00	0.78	0.57	0.29	0.35	0.27	0.23	0.22
96	1.00	1.00	1.00	0.99	0.98	1.00	0.83	0.90	0.75	0.26	0.50	0.53
95	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.88	0.95	0.96	0.95	0.77
94	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.97	1.00	1.00	1.00
93	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.85	1.00	1.00	1.00
92	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.20	1.00	1.00
91	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.16	1.00
90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.11

Table 6: Accuracy of the `BC_threshold` test with $t = t_2$

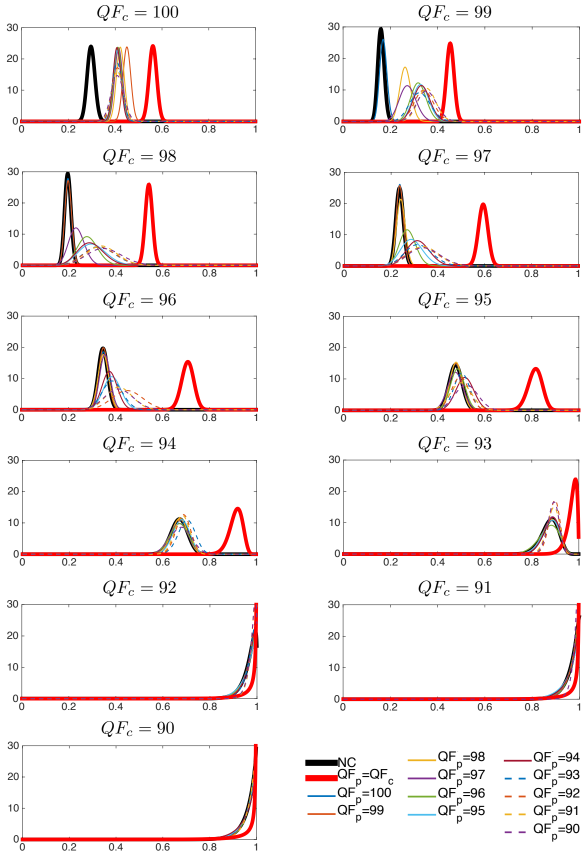


Figure 5: Beta distributions of r fitted for different QF$_c$ and QF$_p$. In each plot the black bold line represents the single compression case with the corresponding QF$_c$, while the red bold line represents the identical double compression [QF$_c$, QF$_c$]; other previous quality factors are reported in the legend.

in the cases of single compression and double compression with QF$_p$ >QF$_c$.

Also in light of the results obtained in Section 4.2, this suggests to reformulate the test with the goal of correctly distinguishing **SC** and **IDC**. In this case, the fitted distribution is clearly separated from the other ones (at least for QF$_c \geq 94$) and represents a relevant open issue for the Benford–Fourier analysis. Moreover, Figure 5 also indicates that for QF$_c \leq 93$ almost all of the blocks are already stable. Thus, we would not gain any information by recompressing the image multiple times and get the whole convergence path, as suggested in [3].

Then, we can design a simple threshold-based test (indicated as `BC_threshold`) on r such that an image is classified as belonging to **SC** ∪ **DDC** if $r \leq t$, or to **IDC** otherwise. The choice of the threshold t can be performed according to different criteria related to the application scenario. As an example, we report in Table 5 and 6 the results obtained by fixing the threshold for each QF$_c$ in two different ways:

- t_1 is such that $\int_0^{t_1} p_{[\mathrm{QF}_c,\mathrm{QF}_c]}(r)dr = 0.01$ (we target 99% accuracy on **IDC**),
- t_2 is such that $\int_{t_2}^{1} p_{[NC,\mathrm{QF}_c]}(r)dr = 0.01$ (we target 99% accuracy on **SC**.

In practice, we have that both thresholds yield good accuracies for **IDC** when QF$_c \geq 94$ (as it can be expected from Fig. 5), while if QF$_c \leq 93$ we have misclassification either for **SC** and **DDC** (with $t = t_1$) or **IDC** (with $t = t_2$). Then, we can consider the threshold-based approach on block convergence ratio as accurate for the discrimination of single compressed and identically recompressed images when quality factors are ≥ 94.

5. COMBINED APPROACH

In light of the results from Section 4, we can notice that the pros and cons of the two techniques are mostly complementary, thus suggesting the development of a combined approach for coping with the HQ-DC scenario. In particular, results from the previous section show that the `BF_adaptive` test distinguishes with good accuracy images in **SC** ∪ **IDC** from images in **DDC** (with misclassification cases for QF$_p$ > QF$_c$); on the other hand, the `BC_threshold` correctly distinguishes images in **SC** ∪ **DDC** from images in **IDC** (with misclassification cases for QF$_c \leq 93$).

The goal is to design a classification test for high quality JPEG images that is able to correctly assign an image to one of the three classes (**SC**, **DDC** or **IDC**) by relying on the knowledge of LLR_m (the maximum LLR value obtained from the Benford–Fourier analysis by excluding QF$_c$) and r,

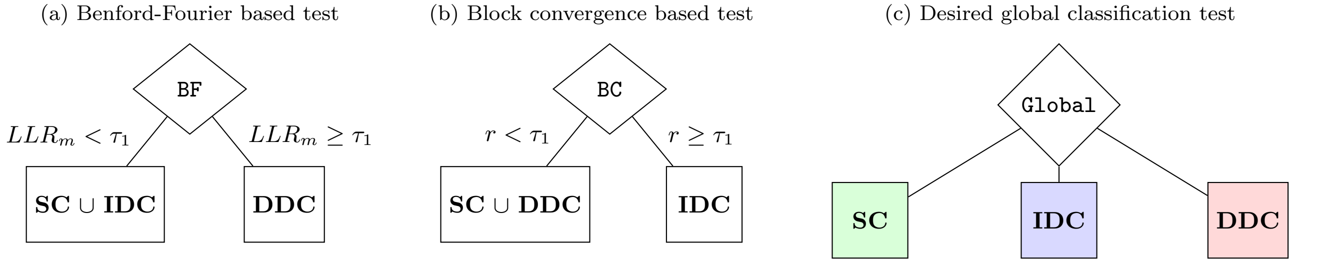


Figure 6: Decision trees for different tests.

the ratio of stable blocks after recompressing with QF_c.

This task can be accomplished by means of decision tree induction theory [2, 14], which allows us to determine decision rules on the pair $(LLR_m, r) \in \mathbb{R} \times [0, 1]$ obtained from the analyzed image.

5.1 Decision tree induction

The decision tree (DT) is one of the most used ways to represent a classification test, which is expressed as a recursive partitioning of the instance space (in our case $\mathbb{R} \times [0, 1]$). The problem of building (or *inducing*) a tree starting from a set of labeled cases (i.e., a set of attribute tuples and corresponding classes) has been extensively studied in the literature [2], providing a number of effective and efficient solutions. A tree is composed by a number of nodes, each of them related to a specific attribute thresholding operation. Nodes that are followed by a subtree are called internal nodes. Otherwise, they are called leaves and they represent the fact that a decision has been reached (i.e, the analyzed attribute tuple has been assigned to a class and no further thresholding is performed). In our case every attribute tuple is a pair (LLR_m, r), while the possible classes are **SC**, **DDC** or **IDC**.

In our visual representations of the trees, we will indicate internal nodes as diamonds containing the name of the test used (i.e., `BF` for Benford–Fourier and `BC` for block convergence analysis), while leaves are denoted with squares. We number each internal node in top-bottom/left-right order and we denoted as τ_i the threshold used in the i-th internal node. As an example, in Fig. 6 we represent the classification tests used in Section 4, and the one that we want to design. Please, note that the global test in Fig. 6c discriminates between the three different classes and the accuracy values in the following tables will be computed accordingly, i.e., reporting in each cell the percentage of images that have been correctly assigned to **SC**, **IDC** or **DDC**.

The induction process generally consists of a *growing* phase, where the tree is developed according to greedy algorithms, and a *pruning* phase, where the tree is further reduced by replacing a subtree with a decision leaf [14]. All these operations pursue the goal of maximizing accuracy on a given training set (a set of samples for which both the attributes and the corresponding classes are known) while minimizing the complexity of the tree, and are performed according to certain criteria and metrics. The result is a list of sequential decision rules that indicates how to optimally threshold the values LLR_m and r, and in which order. In the following, DT induction is used to derive accurate classification tests starting from a number of labeled training images.

It is worth pointing out that, differently from other kind of classifiers, decision trees are easy to interpret and represent. Moreover, their complexity can be controlled in several ways, like fixing a maximum number of nodes or a minimum number of leaves. This results in a classification test combining multiple attributes that can be easily conveyed and explained.

5.2 Experimental set-up

In our test, we consider the 738 UCID images used in the previous section and, additionally, 500 DRESDEN images. For both datasets a number of images has been randomly chosen for the training set of the DT (200 and 100 images, respectively), while the remaining ones are used for testing.

Among the existing toolboxes available for the DT induction, we used the `fitctree` MatLab function contained in the Statistics and Machine Learning Toolbox. We used the default options, with the exception of the prior probability of each class, that we explicitly set uniform. In other words, we consider as equally probable images in **SC**, **DDC** and **IDC**.

Moreover, in each experiment we both determine a *full* DT (i.e., the one that is built in the growing phase) and a *pruned* one, obtained by forcing the algorithm to reduce the tree until it contains less than 8 nodes, in order to have a simplified version.

5.3 Overall decision tree

We first try to build a DT that can be applied to a high quality image regardless of its current quality factor. In this case, the training set is composed of each training image processed according to the 132 different quantization chains, i.e., by using all the quality factors.

The full DTs obtained are quite complex for both datasets, presenting 3137 and 1395 nodes for UCID and DRESDEN datasets, respectively. They are quite accurate on the training set, while the performance strongly degrades when moving to the testing set (see Tables 11, 12, 13 and 14 in Appendix A.1). This suggests that the algorithm is forced to create a high number of nodes to cope with the specificity of the training set, but it is sensitive to the images it contains. In other words, it suffers from overfitting.

On the other hand, it is worth noticing that the pruned DTs obtained (reported in Fig. 7) have the same structure for both datasets (i.e., using BF analysis first to identify **DDC** images and then employ block convergence to distinguish between **SC** and **IDC**), thus differing only in the thresholds used.

Moreover, by observing the accuracies (reported in Tables 7 and 8) we can notice that the pruned versions achieve good results both in the training and testing set when QF_c is high,

(a) Training set

	NC	**100**	**99**	**98**	**97**	**96**	**95**	**94**	**93**	**92**	**91**	**90**
100	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
99	0.97	0.10	0.97	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
98	0.96	0.11	0.15	0.95	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
97	0.95	0.12	0.17	0.69	0.95	1.00	1.00	1.00	1.00	1.00	1.00	1.00
96	0.99	0.04	0.03	0.39	0.79	0.99	0.98	1.00	1.00	1.00	1.00	1.00
95	0.10	0.06	0.06	0.15	0.76	0.91	0.98	0.98	0.99	1.00	1.00	1.00
94	0.01	0.02	0.03	0.04	0.04	0.95	0.99	0.99	0.98	0.99	1.00	0.99
93	0.01	0.01	0.01	0.07	0.12	0.91	0.98	0.98	1.00	0.96	0.98	0.99
92	0.00	0.01	0.01	0.02	0.01	0.35	0.91	0.96	0.97	1.00	0.95	0.99
91	0.00	0.01	0.03	0.43	0.56	0.69	0.97	1.00	1.00	0.96	0.99	0.98
90	0.01	0.00	0.01	0.01	0.06	0.21	0.81	1.00	0.99	0.98	0.93	1.00

(b) Testing set

	NC	**100**	**99**	**98**	**97**	**96**	**95**	**94**	**93**	**92**	**91**	**90**
100	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
99	0.95	0.13	0.95	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
98	0.95	0.13	0.16	0.96	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
97	0.94	0.11	0.16	0.83	0.94	1.00	1.00	1.00	1.00	1.00	1.00	1.00
96	0.98	0.04	0.07	0.41	0.90	0.98	1.00	1.00	1.00	1.00	1.00	1.00
95	0.01	0.03	0.04	0.17	0.87	0.95	0.99	0.99	1.00	1.00	1.00	1.00
94	0.00	0.00	0.02	0.03	0.04	0.98	0.98	1.00	0.97	1.00	1.00	1.00
93	0.00	0.01	0.01	0.15	0.16	0.96	0.99	0.96	0.99	0.96	0.98	1.00
92	0.00	0.01	0.01	0.02	0.03	0.45	0.95	0.97	0.95	1.00	0.96	1.00
91	0.00	0.00	0.01	0.48	0.68	0.83	1.00	1.00	0.99	0.92	1.00	0.95
90	0.00	0.00	0.00	0.01	0.09	0.28	0.85	1.00	1.00	0.97	0.90	1.00

Table 7: Accuracies of overall pruned DT for UCID dataset

(a) Training set

	NC	**100**	**99**	**98**	**97**	**96**	**95**	**94**	**93**	**92**	**91**	**90**
100	1.00	1.00	0.92	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
99	1.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
98	0.99	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
97	1.00	0.00	0.00	0.20	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
96	1.00	0.00	0.00	0.03	0.36	1.00	1.00	1.00	1.00	1.00	1.00	1.00
95	0.00	0.00	0.00	0.03	0.24	0.86	1.00	1.00	1.00	1.00	1.00	1.00
94	0.00	0.00	0.00	0.00	0.00	0.87	1.00	1.00	1.00	1.00	1.00	1.00
93	0.00	0.01	0.01	0.01	0.01	0.66	0.96	1.00	1.00	1.00	1.00	1.00
92	0.00	0.01	0.01	0.01	0.01	0.03	0.83	0.99	1.00	1.00	1.00	1.00
91	0.00	0.01	0.01	0.11	0.14	0.18	0.88	1.00	1.00	1.00	1.00	1.00
90	0.00	0.01	0.01	0.02	0.02	0.02	0.89	0.99	1.00	1.00	1.00	1.00

(b) Testing set

	NC	**100**	**99**	**98**	**97**	**96**	**95**	**94**	**93**	**92**	**91**	**90**
100	1.00	1.00	0.91	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
99	1.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
98	1.00	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
97	1.00	0.00	0.00	0.23	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
96	0.97	0.00	0.00	0.02	0.34	1.00	1.00	1.00	1.00	1.00	1.00	1.00
95	0.00	0.00	0.00	0.03	0.20	0.85	1.00	1.00	1.00	1.00	1.00	1.00
94	0.00	0.00	0.01	0.01	0.01	0.89	1.00	1.00	1.00	1.00	1.00	1.00
93	0.00	0.01	0.02	0.02	0.02	0.62	0.96	1.00	0.99	1.00	1.00	1.00
92	0.00	0.02	0.02	0.02	0.02	0.03	0.84	0.99	1.00	0.99	1.00	1.00
91	0.00	0.02	0.02	0.12	0.18	0.16	0.90	0.99	1.00	1.00	1.00	1.00
90	0.00	0.02	0.03	0.04	0.05	0.05	0.91	0.98	1.00	1.00	1.00	1.00

Table 8: Accuracies of overall pruned DT for DRESDEN dataset

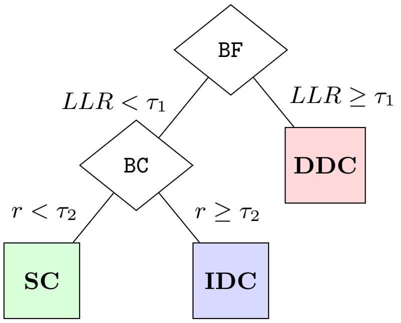


Figure 7: Overall best pruned decision tree

whereas the misclassification for **SC** is noticeably higher for lower values of QF$_c$ in both datasets.

In line with what we observed in Section 4, these results confirm that the thresholds to be used differ when varying QF$_c$ due to the non-homogeneous behavior of the attributes, especially for r.

5.4 QF$_c$-specific decision trees

Given the results obtained when using the same thresholds for every image, a reasonable solution would be to differentiate the classification test according to the current quality factor, i.e., performing the tree induction process separately for different QF$_c$. Indeed, it is worth observing that the current quality factor is known, thus such approach is feasible in a realistic forensic scenario.

We repeat the DT building for the 11 different values of QF$_c$, where the training set is now composed only of images compressed once or twice with QF$_c$ as last quality factor. On the one hand, we have that the full trees have very good accuracies on the training sets but the performance degrades when applied to the testing sets, as shown in Tables 15, 16, 17 and 18 in Appendix A.2.

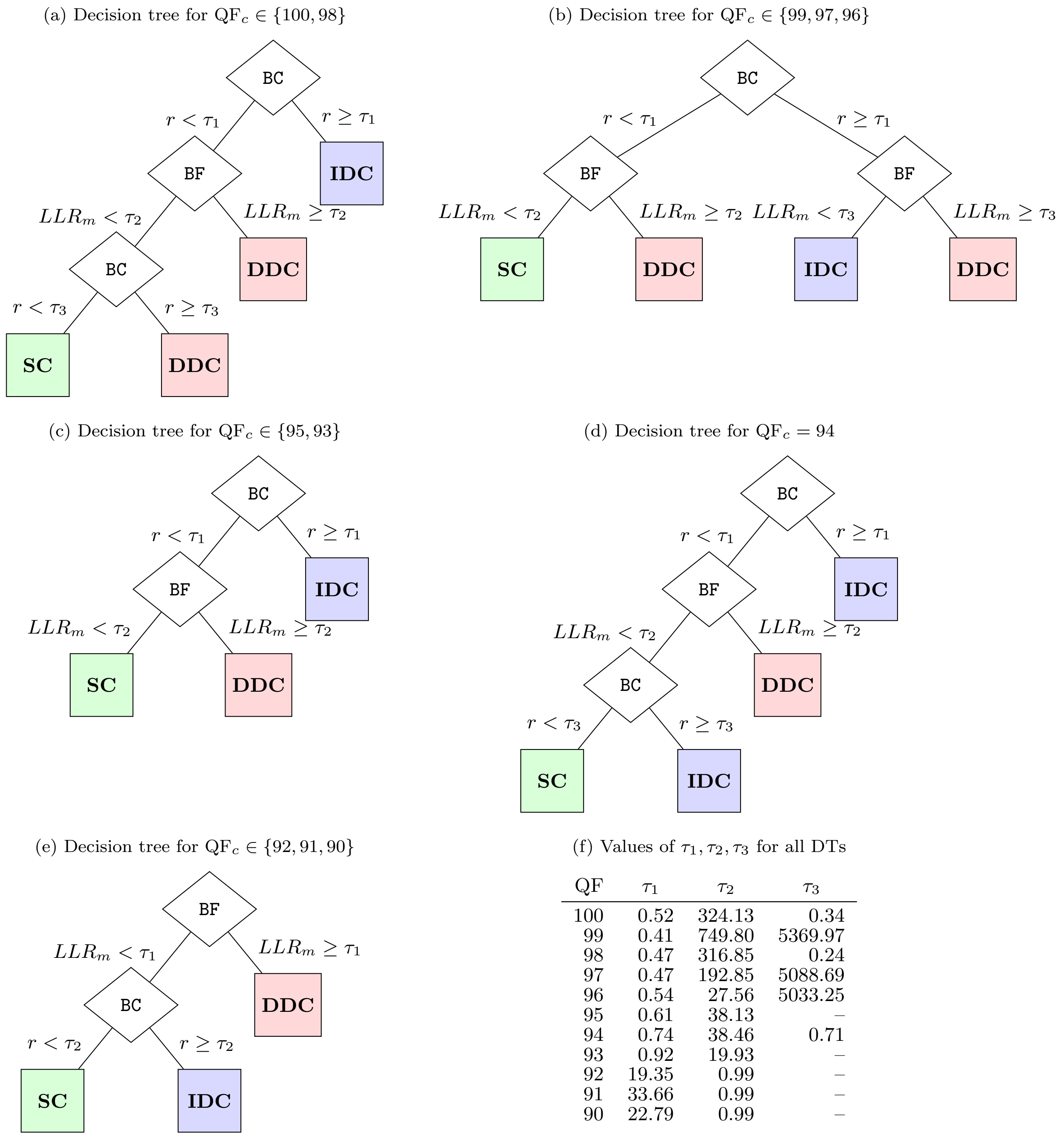
On the other hand, pruned trees lead to stable results for training and testing set for both datasets (Tables 9 and 10). For the sake of brevity, we only report the pruned trees obtained from the UCID dataset (Fig. 8), together with the different thresholds determined in each case. It is interesting to observe how the structure of tree varies among the quality factors, allowing either two or three levels of depth and splitting the nodes in different ways. For instance, we can observe that the first attribute chosen by the algorithm is LLR_m for QF$_c \leq 92$, while it switches to r for higher QF$_c$ for which the block convergence is more accurate.

The results in Tables 9 and 10 indicate that the pruned DTs determined separately for different values of QF$_c$ yield accurate results. Note that in Tables 3, 5 and 6 the accuracies are computed with respect to the classes discriminated in the two single tests (**SC** ∪ **IDC** vs **DDC** for BF and **SC** ∪ **DDC** vs **IDC** for BC), while the QF$_c$-specific DTs achieve good accuracy for all the three possible classes. Thus, we can conclude that the capability of BF and BC of correctly identifying **DDC** and **IDC**, respectively, is generally maintained, while the global misclassification on the three classes is highly reduced.

In light of these results, we consider the QF$_c$-specific decision trees as a possible effective solution for the distinction of images in **SC**, **DDC** and **IDC** in the HQ-DC scenario.

6. CONCLUSIONS

We have addressed the single vs double compression discrimination problem for grayscale JPEG images compressed with high nearly-identical quality factors (≥ 90). After analyzing the performance of the Benford–Fourier analysis in the DCT domain and the block convergence analysis in the



(a) Decision tree for $\mathrm{QF}_c \in \{100, 98\}$

(b) Decision tree for $\mathrm{QF}_c \in \{99, 97, 96\}$

(c) Decision tree for $\mathrm{QF}_c \in \{95, 93\}$

(d) Decision tree for $\mathrm{QF}_c = 94$

(e) Decision tree for $\mathrm{QF}_c \in \{92, 91, 90\}$

(f) Values of τ_1, τ_2, τ_3 for all DTs

QF	τ_1	τ_2	τ_3
100	0.52	324.13	0.34
99	0.41	749.80	5369.97
98	0.47	316.85	0.24
97	0.47	192.85	5088.69
96	0.54	27.56	5033.25
95	0.61	38.13	–
94	0.74	38.46	0.71
93	0.92	19.93	–
92	19.35	0.99	–
91	33.66	0.99	–
90	22.79	0.99	–

Figure 8: Pruned decision trees and thresholds for different quality factors for the UCID dataset

pixel domain, we have studied the problem of combining the two techniques to obtain an accurate discrimination between single compressed images (**SC**), double compressed images with a different quality factor (**DDC**) and images recompressed with the same quality factor (**IDC**).

The final set of detectors on both the datasets considered proves to be very accurate for **SC** images (accuracy $\geq$ 97.5%), **DDC** images with $\mathrm{QF}_p < \mathrm{QF}_c$ (accuracy $\geq$ 99.0%) and **IDC** with $\mathrm{QF}_c \geq 93$ (accuracy $\geq$ 99.5%).

The results obtained suggest a number of open issues and directions for future work. The first evident space of improvement is represented by the low detection rate of certain compression chains. For instance, the **DDC** cases where $\mathrm{QF}_p > \mathrm{QF}_c$ are often misclassified when $\mathrm{QF}_p \geq 96$. The same happens for the **IDC** images when $\mathrm{QF}_c \leq 92$. Indeed, none of the methods considered is able to correctly identify them and the combined global test does not achieve good performance in those cases, although it leads to improvements with respect to the two separate techniques. As a future perspective, additional methods could be used to cope with these specific issues and incorporated in the final decision tree. For instance, the approaches in [6] or [16] could

(a) Training set.

	NC	**100**	**99**	**98**	**97**	**96**	**95**	**94**	**93**	**92**	**91**	**90**
100	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
99	1.00	0.04	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
98	1.00	0.04	0.09	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99
97	1.00	0.01	0.03	0.68	0.99	0.98	1.00	1.00	1.00	1.00	1.00	1.00
96	0.99	0.04	0.04	0.40	0.80	0.99	0.98	1.00	1.00	1.00	1.00	1.00
95	0.99	0.05	0.05	0.15	0.76	0.91	0.99	0.97	0.98	0.99	1.00	0.99
94	1.00	0.02	0.02	0.02	0.04	0.95	0.99	0.98	0.91	0.98	0.99	0.99
93	1.00	0.02	0.06	0.14	0.17	0.93	0.98	0.99	0.86	0.97	0.98	1.00
92	0.99	0.03	0.06	0.07	0.06	0.39	0.92	0.97	0.98	0.36	0.97	1.00
91	0.98	0.01	0.03	0.44	0.56	0.69	0.97	1.00	1.00	0.96	0.28	0.98
90	1.00	0.01	0.02	0.03	0.10	0.24	0.85	1.00	0.99	1.00	0.95	0.23

(b) Testing set.

	NC	**100**	**99**	**98**	**97**	**96**	**95**	**94**	**93**	**92**	**91**	**90**
100	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00
99	0.99	0.08	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
98	0.99	0.02	0.06	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.99	0.98
97	1.00	0.01	0.01	0.79	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
96	0.98	0.06	0.08	0.42	0.91	1.00	0.99	1.00	1.00	1.00	1.00	1.00
95	0.99	0.03	0.04	0.15	0.86	0.95	1.00	0.98	1.00	1.00	1.00	1.00
94	1.00	0.00	0.01	0.02	0.03	0.98	0.98	1.00	0.92	1.00	1.00	1.00
93	0.99	0.03	0.04	0.20	0.20	0.95	0.99	0.98	0.96	0.98	0.99	1.00
92	0.99	0.02	0.04	0.05	0.07	0.49	0.95	0.99	0.97	0.40	0.97	1.00
91	0.99	0.00	0.01	0.48	0.68	0.83	1.00	1.00	0.99	0.92	0.28	0.96
90	0.99	0.01	0.01	0.01	0.14	0.32	0.88	1.00	1.00	0.99	0.94	0.26

Table 9: Accuracies of QF_c-specific pruned DT for UCID dataset.

(a) Training set.

	NC	**100**	**99**	**98**	**97**	**96**	**95**	**94**	**93**	**92**	**91**	**90**
100	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
99	1.00	0.03	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
98	1.00	0.00	0.01	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
97	0.96	0.01	0.04	0.79	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.98
96	1.00	0.02	0.01	0.07	0.88	1.00	1.00	1.00	1.00	1.00	1.00	1.00
95	0.98	0.06	0.10	0.30	0.47	0.93	1.00	1.00	1.00	1.00	1.00	1.00
94	1.00	0.01	0.01	0.01	0.01	0.87	1.00	1.00	1.00	1.00	1.00	1.00
93	0.97	0.10	0.11	0.16	0.13	0.78	0.97	1.00	1.00	1.00	1.00	1.00
92	0.99	0.04	0.07	0.07	0.07	0.08	0.89	0.99	1.00	0.87	0.73	0.99
91	1.00	0.01	0.01	0.11	0.14	0.18	0.88	1.00	1.00	1.00	0.78	0.88
90	0.99	0.04	0.08	0.08	0.10	0.11	0.90	0.99	1.00	1.00	1.00	0.76

(b) Testing set.

	NC	**100**	**99**	**98**	**97**	**96**	**95**	**94**	**93**	**92**	**91**	**90**
100	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
99	0.99	0.02	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
98	1.00	0.00	0.01	1.00	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00
97	0.94	0.04	0.05	0.77	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99
96	0.97	0.04	0.04	0.06	0.85	1.00	0.99	0.99	0.99	1.00	0.99	0.99
95	0.97	0.06	0.07	0.28	0.36	0.95	1.00	1.00	1.00	1.00	1.00	1.00
94	0.99	0.01	0.02	0.02	0.02	0.89	1.00	0.99	1.00	1.00	1.00	1.00
93	0.94	0.11	0.12	0.14	0.13	0.71	0.97	1.00	0.99	1.00	1.00	1.00
92	0.96	0.07	0.07	0.09	0.08	0.09	0.86	0.99	1.00	0.89	0.68	0.98
91	1.00	0.02	0.02	0.12	0.18	0.16	0.90	0.99	1.00	1.00	0.78	0.85
90	0.97	0.07	0.09	0.10	0.11	0.11	0.92	0.98	1.00	0.99	0.99	0.75

Table 10: Accuracies of QF_c-specific pruned DT for DRESDEN dataset.

be employed for identifying **IDC** with $QF_c \leq 92$. Moreover, a limitation of the proposed approach is that it does not explicitly incorporate the knowledge of the size of the image under investigation.

Finally, the process of decision tree induction is currently performed by means of standard tools. A potential improvement would be to design induction tools specifically tailored to the forensic scenario considered, by customizing the criteria that rule the construction of the tree.

7. REFERENCES

[1] T. Bianchi, A. Piva, and F. Pérez-González. Near optimal detection of quantized signals and application to JPEG forensics. In *IEEE Workshop on Informations Forensics and Security (WIFS)*, 2013.

[2] L. Breslow and D. Aha. Simplifying decision trees: A survey. *The Knowledge Engineering Review*, 12(1):1–40, 1997.

[3] M. Carnein, P. Schöttle, and R. Böhme. Forensics of high-quality JPEG images with color subsampling. In *IEEE Workshop on Informations Forensics and Security (WIFS)*, 2015.

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

[5] T. Gloe and R. Boehme. The Dresden image database for benchmarking digital image forensics. In *ACM Symposium on Applied Computing*, volume 2, pages 1585–1591, 2010.

[6] F. Huang, J. Huang, and Y. Shi. Detecting double JPEG compression with the same quantization matrix. *IEEE Transactions on Information Forensics and Security*, 5(4):848–856, 2010.

[7] S.-Y. Lai and R. 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)*, pages 3028–3032, 2013.

[8] B. Li, Y. Shi, and J. Huang. Detecting doubly compressed JPEG images by using mode based first digit features. In *IEEE Workshop on Multimedia Signal Processing (MMSP)*, pages 730–735, 2008.

[9] S. Milani, M. Tagliasacchi, and S. Tubaro. Discriminating multiple JPEG compression using first digit features. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 25–30, 2012.

[10] C. Pasquini, G. Boato, and F. Pérez-González. Multiple JPEG compression detection by means of Benford-Fourier coefficients. In *IEEE Workshop on Informations Forensics and Security (WIFS)*, pages 113–118, 2014.

[11] C. Pasquini, F. Pérez-González, and G. Boato. A Benford-Fourier JPEG compression detector. In *IEEE International Conference on Image Processing (ICIP)*, pages 5322–5326, 2014.

[12] 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.

[13] T. Pevný and J. Fridrich. Detection of double-compression in JPEG images for applications in steganography. *IEEE Transactions on Information Forensics and Security*, 3(2):247–258, 2008.

[14] L. Rokach and O. Maimon. Top-down induction of decision trees classifiers - a survey. *IEEE Transactions on Systems, Man, and Cybernetics*, 35(4):476–487,

2005.

[15] G. Schaefer and M. Stich. UCID - An uncompressed colour image database. In *SPIE Storage and Retrieval Methods and Applications to Multimedia*, volume 5307, 2004.

[16] J. Yang, J. Xie, G. Zhu, S. Kwong, and Y. Shi. An effective method for detecting double JPEG compression with the same quantization matrix. *IEEE Transactions on Information Forensics and Security*, 9(11), 2014.

APPENDIX

A. RESULTS FOR FULL DECISION TREES

A.1 Overall full decision tree

	NC	**100**	**99**	**98**	**97**	**96**	**95**	**94**	**93**	**92**	**91**	**90**
100	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
99	1.00	0.62	0.96	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
98	1.00	0.67	0.69	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
97	1.00	0.64	0.64	0.85	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
96	1.00	0.70	0.72	0.79	0.93	0.96	0.99	1.00	1.00	0.99	1.00	1.00
95	0.98	0.64	0.63	0.70	0.91	0.96	0.99	0.99	1.00	1.00	1.00	1.00
94	0.96	0.59	0.60	0.61	0.64	0.95	0.98	0.94	0.98	0.99	1.00	0.99
93	0.95	0.49	0.53	0.59	0.60	0.94	0.98	0.99	0.92	0.98	0.99	0.99
92	0.95	0.50	0.53	0.54	0.54	0.69	0.95	0.99	0.99	0.93	0.99	1.00
91	0.97	0.49	0.55	0.76	0.83	0.85	0.99	1.00	1.00	1.00	0.95	0.99
90	0.94	0.50	0.56	0.55	0.58	0.65	0.92	0.99	0.99	0.99	0.97	0.96

Table 11: Training set UCID

	NC	**100**	**99**	**98**	**97**	**96**	**95**	**94**	**93**	**92**	**91**	**90**
100	0.91	0.95	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
99	0.70	0.44	0.63	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
98	0.54	0.45	0.46	0.93	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
97	0.54	0.49	0.47	0.89	0.86	1.00	1.00	1.00	1.00	1.00	1.00	1.00
96	0.45	0.61	0.60	0.75	0.96	0.67	1.00	1.00	1.00	1.00	1.00	1.00
95	0.35	0.56	0.58	0.66	0.93	0.94	0.77	0.98	0.99	1.00	1.00	1.00
94	0.36	0.48	0.48	0.51	0.49	0.98	0.98	0.57	0.98	1.00	1.00	1.00
93	0.39	0.36	0.32	0.46	0.47	0.97	0.99	0.98	0.45	0.99	0.99	1.00
92	0.32	0.34	0.40	0.37	0.40	0.64	0.97	0.99	0.98	0.58	0.99	1.00
91	0.32	0.34	0.42	0.72	0.84	0.89	1.00	1.00	1.00	0.96	0.56	0.99
90	0.33	0.36	0.39	0.43	0.52	0.58	0.94	1.00	1.00	0.99	0.96	0.53

Table 12: Testing set UCID

	NC	**100**	**99**	**98**	**97**	**96**	**95**	**94**	**93**	**92**	**91**	**90**
100	1.00	1.00	0.96	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
99	1.00	0.68	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
98	0.99	0.68	0.66	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
97	1.00	0.68	0.63	0.82	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
96	1.00	0.74	0.78	0.75	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00
95	0.96	0.71	0.67	0.80	0.79	0.98	1.00	1.00	1.00	1.00	1.00	1.00
94	0.99	0.81	0.78	0.81	0.86	0.99	1.00	0.99	1.00	1.00	1.00	1.00
93	1.00	0.71	0.75	0.79	0.85	0.94	1.00	1.00	0.94	1.00	1.00	1.00
92	0.94	0.66	0.71	0.71	0.70	0.64	0.96	1.00	1.00	1.00	1.00	1.00
91	1.00	0.58	0.59	0.69	0.73	0.73	0.93	1.00	1.00	1.00	0.99	1.00
90	0.99	0.76	0.77	0.82	0.74	0.75	0.96	0.99	1.00	1.00	1.00	0.98

Table 13: Training set DRESDEN

	NC	**100**	**99**	**98**	**97**	**96**	**95**	**94**	**93**	**92**	**91**	**90**
100	0.93	0.98	0.93	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
99	0.74	0.47	0.93	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
98	0.54	0.54	0.45	0.96	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
97	0.56	0.42	0.43	0.68	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00
96	0.54	0.62	0.64	0.66	0.97	0.97	1.00	1.00	1.00	1.00	1.00	1.00
95	0.53	0.49	0.51	0.71	0.76	0.96	0.99	1.00	1.00	1.00	1.00	1.00
94	0.43	0.66	0.69	0.67	0.63	0.97	1.00	0.96	1.00	1.00	1.00	1.00
93	0.43	0.63	0.67	0.63	0.66	0.91	0.98	1.00	0.51	1.00	1.00	1.00
92	0.40	0.45	0.49	0.48	0.49	0.46	0.93	1.00	1.00	0.88	1.00	1.00
91	0.42	0.49	0.51	0.55	0.59	0.60	0.93	1.00	1.00	1.00	0.87	1.00
90	0.34	0.52	0.54	0.60	0.58	0.58	0.95	0.99	1.00	1.00	1.00	0.83

Table 14: Testing set DRESDEN

A.2 QF_c-specific full decision trees

	NC	**100**	**99**	**98**	**97**	**96**	**95**	**94**	**93**	**92**	**91**	**90**
100	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
99	1.00	0.61	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
98	1.00	0.65	0.66	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
97	1.00	0.63	0.67	0.86	1.00	0.98	1.00	1.00	1.00	1.00	1.00	1.00
96	1.00	0.78	0.69	0.82	0.94	1.00	0.99	1.00	1.00	1.00	1.00	1.00
95	1.00	0.73	0.73	0.81	0.95	0.98	0.99	1.00	1.00	1.00	1.00	1.00
94	1.00	0.77	0.80	0.74	0.79	0.98	1.00	0.99	0.99	0.99	1.00	1.00
93	0.98	0.66	0.73	0.74	0.73	0.98	1.00	1.00	0.98	0.98	0.99	1.00
92	0.97	0.48	0.61	0.60	0.58	0.75	0.96	0.99	0.98	0.97	0.99	1.00
91	0.95	0.47	0.53	0.76	0.79	0.88	0.99	1.00	1.00	0.99	0.94	1.00
90	0.95	0.52	0.51	0.68	0.73	0.71	0.96	1.00	1.00	1.00	0.99	0.95

Table 15: Training set UCID

	NC	**100**	**99**	**98**	**97**	**96**	**95**	**94**	**93**	**92**	**91**	**90**
100	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00
99	0.77	0.43	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
98	0.58	0.43	0.47	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
97	0.52	0.49	0.54	0.91	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
96	0.41	0.61	0.63	0.74	0.95	1.00	1.00	1.00	1.00	1.00	1.00	1.00
95	0.43	0.62	0.64	0.71	0.95	0.98	1.00	0.99	1.00	1.00	1.00	1.00
94	0.38	0.64	0.63	0.68	0.64	0.99	0.99	1.00	0.99	1.00	1.00	1.00
93	0.42	0.57	0.58	0.65	0.67	0.98	1.00	0.99	0.97	0.99	0.99	1.00
92	0.40	0.38	0.44	0.45	0.45	0.69	0.97	1.00	0.98	0.55	0.99	1.00
91	0.38	0.32	0.33	0.73	0.80	0.89	1.00	1.00	1.00	0.95	0.56	0.97
90	0.30	0.44	0.51	0.50	0.66	0.66	0.96	1.00	1.00	1.00	0.99	0.46

Table 16: Testing set UCID

	NC	**100**	**99**	**98**	**97**	**96**	**95**	**94**	**93**	**92**	**91**	**90**
100	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
99	1.00	0.61	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
98	1.00	0.51	0.54	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
97	1.00	0.57	0.65	0.88	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
96	1.00	0.71	0.76	0.77	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
95	1.00	0.78	0.74	0.82	0.91	1.00	1.00	1.00	1.00	1.00	1.00	1.00
94	1.00	0.79	0.72	0.83	0.82	0.99	1.00	1.00	1.00	1.00	1.00	1.00
93	1.00	0.77	0.78	0.86	0.77	0.96	0.99	1.00	1.00	1.00	1.00	1.00
92	1.00	0.75	0.79	0.79	0.79	0.76	0.98	0.99	1.00	0.96	1.00	1.00
91	1.00	0.69	0.78	0.79	0.76	0.76	0.97	1.00	1.00	1.00	0.97	1.00
90	0.99	0.62	0.71	0.73	0.75	0.73	0.93	0.99	1.00	1.00	1.00	0.99

Table 17: Training set DRESDEN

	NC	**100**	**99**	**98**	**97**	**96**	**95**	**94**	**93**	**92**	**91**	**90**
100	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
99	0.79	0.40	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
98	0.59	0.43	0.41	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
97	0.57	0.40	0.41	0.84	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
96	0.55	0.62	0.63	0.65	0.97	1.00	1.00	1.00	1.00	1.00	1.00	1.00
95	0.54	0.55	0.57	0.75	0.81	1.00	1.00	1.00	1.00	1.00	1.00	1.00
94	0.51	0.61	0.69	0.72	0.64	0.99	1.00	0.99	1.00	1.00	1.00	1.00
93	0.44	0.60	0.65	0.70	0.65	0.89	0.99	1.00	0.99	1.00	1.00	1.00
92	0.47	0.56	0.60	0.60	0.62	0.65	0.96	1.00	1.00	0.90	1.00	1.00
91	0.41	0.56	0.56	0.57	0.61	0.62	0.94	1.00	1.00	1.00	0.86	1.00
90	0.47	0.49	0.52	0.52	0.56	0.54	0.97	0.99	1.00	1.00	1.00	0.82

Table 18: Testing set DRESDEN