Proceedings of SPIE Vol. 3971:
Security and Watermarking
of Multimedia Contents II
San Jose, Ca., Jan. 2000.

Quantization Watermarking

Joachim J. Eggers
Telecommunications Laboratory
University of Erlangen-Nuremberg
Cauerstrasse 7/NT, 91058 Erlangen, Germany

Bernd Girod
Information Systems Laboratory
Stanford University
Stanford, CA 94305-9510, USA

ABSTRACT

A watermarking scheme for distinguishing different copies of the same multimedia document (fingerprinting) is investigated. Efficient transmission of natural data requires lossy compression, which might impair the embedded watermark. We investigate whether the quantization step in compression schemes can be replaced by dithered quantization to combine fingerprinting and compression.

Dithered quantization offers the possibility of producing perceptually equivalent signals that are not exactly equal. The non-subtractive quantization error can be used as the watermark. We denote the proposed watermarking scheme as "quantization watermarking". Such a scheme is only practical for watermarking applications where the original signal is available to the detector. We analyze the influence of the dither signal on the perceptual quality of the watermarked document and the watermark detection robustness.

Further, the cross-talk between the non-subtractive quantization errors for two different dither realizations is investigated. An analytical description for fine quantization and experimental results for coarse quantization show how the cross-talk depends on the characteristics of the dither signal. The derived properties of quantization watermarking are verified for combined JPEG compression and fingerprinting. The detection robustness for the proposed quantization error watermark is compared with that of an independent additive watermark.

Keywords: Digital watermarking; Dithered quantization; Image watermarking; JPEG

1. INTRODUCTION

Digital watermarking refers to the embedding of information into multimedia data without introducing perceptual changes to the original data. This technique has been proposed to combat the emerging problem of illegal copying of digital audio, image and video data. The watermarks can be used to embed copyright information into the data itself, or to mark different copies of one document, which allows tracing of an illegal copy back to its owner. The latter application, called fingerprinting, is considered in this article. We focus on fingerprinting of still image data.

In contrast to many other watermarking applications, the original signal is usually available to the fingerprint detector, which simplifies the information transmission. On the other hand, fingerprints can suffer from *collusion attacks*, where differently fingerprinted copies of one document are mixed. Boneh and Shaw[1] have constructed codes that are secure against such collusion attacks. However, their scheme assumes that no further attacks against the fingerprints occur. Guth and Pfitzmann[2] extended Boneh and Shaw's scheme to the case where single fingerprint bits are detectable only with a certain bit error rate. In this paper, we do not consider collusion attacks, but we determine the detection error rates for our proposed fingeprinting scheme, which is important for designing the collusion-secure codes.

Natural image data usually provide some room for hidden fingerprints. However, the data are often compressed by lossy algorithms for efficient transmission and storage. This operation can degrade the robustness of embedded fingerprints significantly. Therefore, a method for combined fingerprinting and compression is discussed in this article. Further impairment of embedded fingerprints can occur due to geometrical distortions of the image. This type of attack can decrease the watermark detection performance severely if no proper resynchronization of the watermark detector is implemented. In this paper, we assume synchronized access to the watermarked data. This is not too restrictive, since the availability of the original signal to the detector should enable sufficiently accurate resynchronization.

For a general fingerprinting scheme, as described in Fig. 1, the embedding process can be described by

$$\vec{s}_k = \vec{x} + \vec{w}_k\,, \tag{1}$$

Further author information: Send correspondence to J. J. Eggers. Email: eggers@LNT.de. At the times of submission and acceptance of this paper, B. Girod was with the Telecommunications Laboratory, Univ. Erlangen-Nuremberg; he is now with the Information Systems Laboratory, Stanford Univ.

where $\vec{x}$ denotes the original signal, $\vec{w}_k$ the signal modification by the embedding process, and $\vec{s}_k$ the published signal (fingerprinted copy). $\vec{x}$ is also called the "host signal". For a secure fingerprinting scheme, a key is mandatory. For instance, the seed of the involved pseudo-random sequence generator generating the watermarks can serve as a key. Some watermarking schemes, including the one proposed in this article, are more complex than the simple addition of a sequence. However, in all cases the watermark can be defined to be the difference between the original signal and watermarked copy, and thus (1) is a valid generalization. In the remainder of this article, signals are denoted by vectors (e.g., $\vec{x}$), the nth signal sample of $\vec{x}$ by $x[n]$, where $n \in \mathcal{N}$ and $\mathcal{N}$ is the set of sample indices, and random variables by boldface (e.g.,$\mathbf{x}$). The index k denotes the fingerprint number.

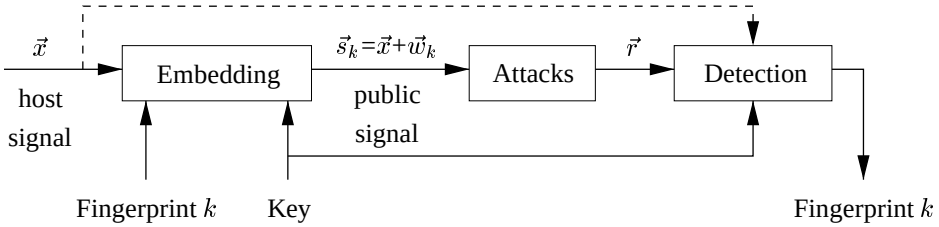

Figure 1. Model of fingerprinting

The fingerprint detector receives a signal $\vec{r}_k = \vec{s}_k + \vec{v} = \vec{x} + \vec{w}_k + \vec{v}$, where $\vec{v}$ includes any distortion that might be introduced between fingerprint embedding and detection. We mentioned already that lossy compression is one of the operations introducing distortion. However, in many applications compression might occur right after fingerprint embedding. Thus, combined fingerprinting and compression seems to be reasonable. The purpose of this article is to investigate how much can be gained by doing so.

A crucial component in lossy compression is (coarse) quantization of the signal samples. Dithered quantization allows one to modify the quantization process, usually without significant changes to the document quality. Thus, we propose combining fingerprinting and compression by replacing the common quantization step in compression algorithms with dithered quantizers. In particular, we consider uniform scalar quantization as used in JPEG compression of still images. We denote this approach by the term **quantization watermarking**. The non-subtractive quantization error is considered the watermark $\vec{w}_k$. Note that this approach is not practical when the host signal is not available to the fingerprint detector.

Dithered uniform scalar quantization has already been used in several proposals for digital watermarking.[3–5] In all these proposals, dithered quantization is combined with other methods to improve watermark detection when the host signal is not available to the detector. From the information-theoretic point of view, the watermark capacity does not depend on the availability of the host signal to the watermark detector.[6,7] Using dithered quantization, it is possible to replace the random codebook in Costa's proof[6] by a systematic one. However, in our case the original is available to the fingerprint detector. Here, dithered quantization is simply used to combine fingerprinting and compression.

Dithered quantization is briefly reviewed in Section 2, and the actual influence of dithered uniform scalar quantization on the signal distortion is discussed. In Section 3 the cross-correlation between quantization errors for different dither sequences is analyzed. Possible detection methods are discussed in Section 4. In Section 5, experimental results for an image fingerprinting scheme are presented, and Section 6 concludes the article.

2. DOCUMENT QUALITY AFTER DITHERED QUANTIZATION

Watermarking of a signal and subsequent quantization can be considered as dithered quantization of the original signal. Dithered quantization has been investigated in the past[8–10] for several purposes. In this section, some important results are reviewed and their relation to watermarking is shown. We consider only uniform scalar quantization and independent identically distributed (I.I.D.) signals.

2.1. Fundamentals of Dithered Quantization

Dithered quantization is an operation in which a signal $\vec{d}$, called **dither**, is added to an input signal $\vec{x}$ prior to quantization. There are two kinds of dithered quantizers, the subtractive dithered quantizer as depicted in Fig. 2(left) , and the non-subtractive

dithered quantizer, as in Fig. 2(right). Subtractive dithered quantization is only realizable when the receiver of the digital data has synchronized access to the dither sequence. The dithered quantization error (DQE) signals for both operations are:

$$\text{subtractive DQE} \quad : \quad \vec{e} = \underline{\vec{s}} - \vec{x} = \vec{s} - \vec{z} \tag{2}$$

$$\text{non-subtractive DQE} \quad : \quad \vec{\epsilon} = \vec{s} - \vec{x} = \vec{e} + \vec{d}, \tag{3}$$

where $\vec{s}$ denotes the output of the non-subtractive dithered quantizer, and $\underline{\vec{s}}$ denotes the output of the subtractive dithered quantizer.

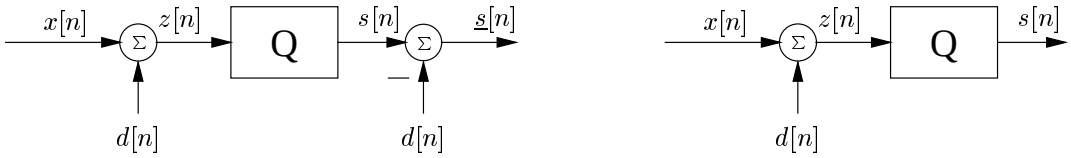

Figure 2. (Left) Subtractive dithered quantization; (right) non-subtractive dithered quantization

The distributor of the signal $\vec{x}$ will send the signal $\vec{s}_k$ to his client k in the case of combined fingerprinting and compression. Thus, the non-subtractive DQE $\vec{\epsilon}_k$ describes the signal distortions due to the embedding process. However, in some cases more insights can be gained by looking at the subtractive DQE $\vec{e}_k$ due to the relationship $\vec{\epsilon}_k = \vec{e}_k + \vec{d}_k$. For brevity, the fingerprint index k is omitted in the remainder of this section.

The previous work on dithered quantization was mainly motivated by the goal of achieving a quantization error that is independent from the quantizer input. This is important, for instance, in the case of universal quantization[11] or to avoid subjectively unpleasant error patterns, e.g., for sinusoidal input sequences. Schuchman[10] showed that the subtractive DQE $\vec{e}$ does not depend on the quantizer input when the dither signal $\vec{d}$ has a uniform distribution within the range of one quantization bin ($d \in [-\Delta/2, \Delta/2]$), leading to an expected squared error of $\mathrm{E}\left\{\mathbf{e}^2\right\} = \Delta^2/12$. Gray and Stockham[8] reviewed and extended this work. They derived dither signals leading to *non-subtractive* DQEs $\vec{\epsilon}$ in which the mth moment does not depend on the quantizer input. Unfortunately, the independence is usually achieved only by increasing the expected squared error $\mathrm{E}\left\{\epsilon^2\right\}$ of the non-subtractive DQE.

2.2. Theoretical Analysis of Quantization Noise Power

The theory on dithered quantization was exploited in[12] to analyze the robustness of common additive watermarking schemes against quantization attacks. This work also enables us to predict the power of the quantization noise for different types of dither signals, and thus, to investigate the embedding distortion of the proposed quantization watermarking scheme. The dither signal $\vec{d}$ is modeled by an I.I.D. random process $\mathbf{d}$ with even symmetric PDF $p_{\mathbf{d}}(d)$ and zero mean $\mathrm{E}\{\mathbf{d}\}$=0. The host signal $\vec{x}$ is modeled by an I.I.D. random process $\mathbf{x}$ with PDF $p_{\mathbf{x}}(x)$. The dither $\mathbf{d}$ is independent from the host $\mathbf{x}$. Therefore, it is convenient to express the statistics of the quantization error in terms of the characteristic functions of the dither signal and the host signal. We define the integral

$$M_{\mathbf{x}}^{(k)}(ju) = \int_{-\infty}^{\infty} x^k \, p_{\mathbf{x}}(x) \exp(jux) \, \mathrm{d}x, \tag{4}$$

which equals the kth derivative of the characteristic function, except for a complex factor. For convenience, the PDFs are normalized by their standard deviations, leading to

$$p_{\tilde{\mathbf{x}}}(x) = \sigma_{\mathbf{x}} \, p_{\mathbf{x}}(\sigma_{\mathbf{x}} x) \tag{5}$$

$$M_{\tilde{\mathbf{x}}}^{(k)}(ju) = \frac{1}{\sigma_{\mathbf{x}}^k} M_{\mathbf{x}}^{(k)}(ju/\sigma_{\mathbf{x}}), \tag{6}$$

where $\tilde{\mathbf{x}}$ indicates the usage of the normalized random variable. In some cases it is useful to relate the standard deviations $\sigma_{\mathbf{x}}$ and $\sigma_{\mathbf{d}}$ to the quantizer step size Δ. We define the normalized parameter $\zeta = \frac{\sigma_{\mathbf{d}}}{\Delta}$ and $\chi = \frac{\sigma_{\mathbf{x}}}{\Delta}$.

Using (3), the distortion introduced by a non-subtractive dithered quantizer can be written as

$$\mathrm{E}\left\{\epsilon^2\right\} = \mathrm{E}\left\{\mathbf{e}^2\right\} + \mathrm{E}\left\{\mathbf{d}^2\right\} + 2\,\mathrm{E}\left\{\mathbf{ed}\right\}. \tag{7}$$

We showed[12] that $\mathrm{E}\{\mathbf{e}^2\}$ and $\mathrm{E}\{\mathbf{ed}\}$ can be computed by

$$\frac{\mathrm{E}\{\mathbf{e}^2\}}{\Delta^2/12} = 1 + 12\sum_{b=1}^{\infty}\frac{(-1)^b}{\pi^2 b^2} M_{\tilde{\mathbf{x}}}(j2\pi b\chi)\, M_{\tilde{\mathbf{d}}}(j2\pi b\zeta) \tag{8}$$

$$\frac{\mathrm{E}\{\mathbf{ed}\}}{\sigma_{\mathbf{d}}^2} = \sum_{b=1}^{\infty}\frac{(-1)^b}{\pi b\zeta} M_{\tilde{\mathbf{x}}}(j2\pi b\chi)\, \mathcal{I}m\left\{M_{\tilde{\mathbf{d}}}^{(1)}(j2\pi b\zeta)\right\}. \tag{9}$$

In our experimental fingerprinting scheme, described in Section 5, the coefficients of an 8×8-block-DCT are watermarked. When these coefficients are modeled by a generalized Gaussian random variable,[12] (8) and (9) can be used to predict the quantization noise power very accurately. For understanding the main influence of the dither signal on the quantization error, it is sufficient to model the host signal by a Laplacian random variable, which equals a generalized Gaussian with shape factor $\nu = 1$. Therefore, only a Laplacian host signal with zero mean will be considered in the remainder of this section. The characteristic function of the Laplacian random variable is $M_{\tilde{\mathbf{x}}}(ju) = \frac{1}{1+\frac{1}{2}u^2}$, and the expected squared quantization error in the undithered case ($\mathbf{d}$=0) is given by

$$\frac{\mathrm{E}\{\epsilon_{\mathbf{d}=0}^2\}}{\Delta^2/12} = \frac{\mathrm{E}\{\mathbf{e}_{\mathbf{d}=0}^2\}}{\Delta^2/12} = 12\chi^2\left(1 - \frac{1}{\sqrt{2}\chi\sinh\frac{1}{\sqrt{2}\chi}}\right). \tag{10}$$

2.3. Embedding Distortion Dependent on Dither Statistics

The influence of the dither distribution on the quantization noise power for different quantizer step sizes Δ is very important for proper design of a quantization watermarking scheme. The power of the non-subtractive DQE $\vec{\epsilon}$ can serve as a measurement for the fingerprint embedding distortion and, thus, should be below the threshold of perceptibility. We discuss results for uniform, Gaussian and bipolar dither distributions.

In Fig. 3, the normalized expected power of the subtractive DQE as a function of χ is shown for $0 < \chi < 1$ and two different values of ζ. Analytic and experimental results are depicted. The excellent agreement between simulated and analytic results confirms the validity of (8). Further, the curves demonstrate the large influence of the dither distribution on the power of the subtractive quantization noise for small values of $\chi = \sigma_{\mathbf{x}}/\Delta$, meaning coarse quantization. For coarse quantization, the subtractive dither quantization noise is much larger than the noise of undithered quantization. On the other hand, neither the dither distribution nor its power has a significant influence on the subtractive quantization noise power for sufficiently fine quantization (large χ). In Fig. 3(left), the power of the dither is chosen such that $\zeta = \sigma_{\mathbf{d}}/\Delta = 1/\sqrt{12}$. For this value of ζ, the uniform dither meets Schuchman's condition, ensuring that the subtractive DQE power is $\mathrm{E}\{\mathbf{e}_{\mathbf{d}\,\mathtt{uniform}}^2\} = \Delta^2/12$, independent of the quantizer input. This effect is clearly visible in Fig. 3(left). We also observe that for all χ the Gaussian dither leads to lower quantization noise than the uniform or the bipolar dither. Similar results for $\zeta = 0.25$ are depicted in Fig. 3(right). For this value of ζ, the bipolar dither has the samples $\pm\Delta/4$, which is a very popular dither sequence in watermarking schemes. Although for all dither signals the quantization noise is monotonically decreasing for an increasing quantizer step size, the noise power is still very close to $\Delta^2/12$ even for very large quantizer step sizes.

Fig. 4(left) depicts the normalized noise power of the non-subtractive DQE, which is calculated using (7), (8) and (9). Since $\zeta = 0.289$, the dither signal has a power of $\Delta^2/12$. For fine quantization, the power of the non-subtractive DQE is approximately the sum of the dither power $\Delta^2/12$ and the power of the subtractive DQE, shown in Fig. 3(left). In this case, the subtractive DQE is almost independent of the dither signal. However, for coarse quantization, the term $\mathrm{E}\{\mathbf{ed}\}$ in (7) becomes important. The non-subtractive DQE $\vec{\epsilon}$ is largest for a Gaussian dither signal for all values of χ. In contrast to Fig. 3, the bipolar dither leads to the smallest non-subtractive quantization noise of all three considered dither distributions. This is due to the stronger dependencies between the bipolar dither and its subtractive quantization noise. Another interesting effect visible in Fig. 4(left) is that the normalized non-subtractive quantization noise tends to zero for very large step size Δ and bipolar or uniform dither signals, whereas a non-zero normalized noise power remains in the case of a Gaussian dither signal. This effect is due to the bounded amplitude of the uniform and bipolar dither in contrast to the unbounded amplitude of Gaussian dither signals.

Finally, Fig. 4(right) depicts the embedding distortion of a quantization watermarking scheme in terms of the document-to-noise ratio (DNR $= 20\log_{10}\sigma_{\mathbf{x}}/\sigma_{\epsilon}$) for different quantizer step sizes and different dither distributions. This plot simply presents the results of Fig. 4(left) in a different way. The triangles show the DNR that would be predicted by the quantization noise formula $\Delta^2/12$, which is only valid for fine quantization. For fine quantization the undithered quantizer achieves about

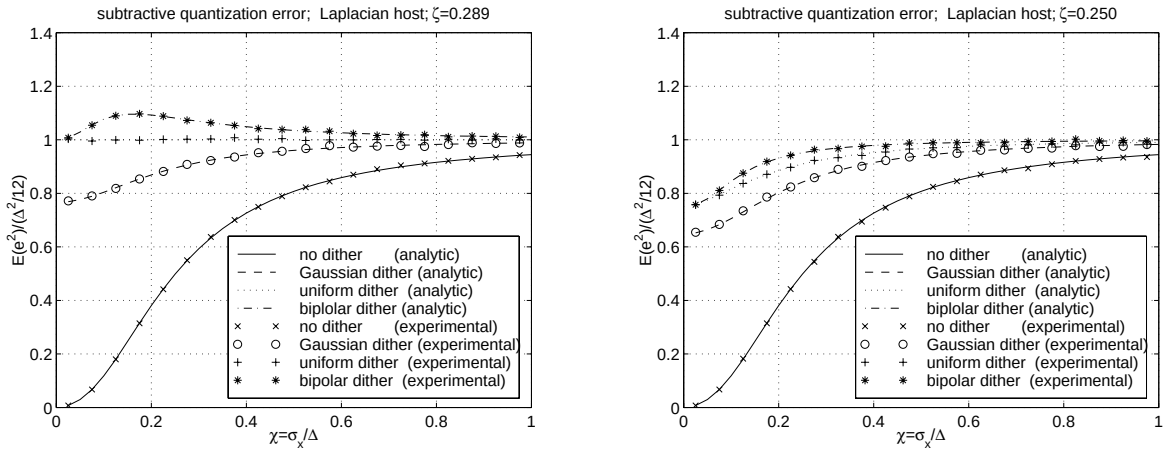

Figure 3. Normalized power of the subtractive DQE $\mathbf{e}$ for different dither distributions and different dither standard deviations; (left) $\zeta = \sigma_{\mathbf{d}}/\Delta = 1/\sqrt{12} = 0.289$; (right) $\zeta = \sigma_{\mathbf{d}}/\Delta = 0.25$

3 dB higher DNR than the different dithered quantizers. For coarse quantization this gap is even larger, except for very coarse quantization where dither-dependent bounding effects occur.

Practical watermarking schemes will only operate at high DNRs, where many of the previously discussed effects do not occur. Fig. 4(right) shows that for high DNR the dither distribution has no significant influence and the embedding distortion can be calculated by the sum of dither power and undithered quantization noise power. The presented theory can be used to determine the compression strength at which it is no longer useful to replace fixed quantization by dithered quantization.

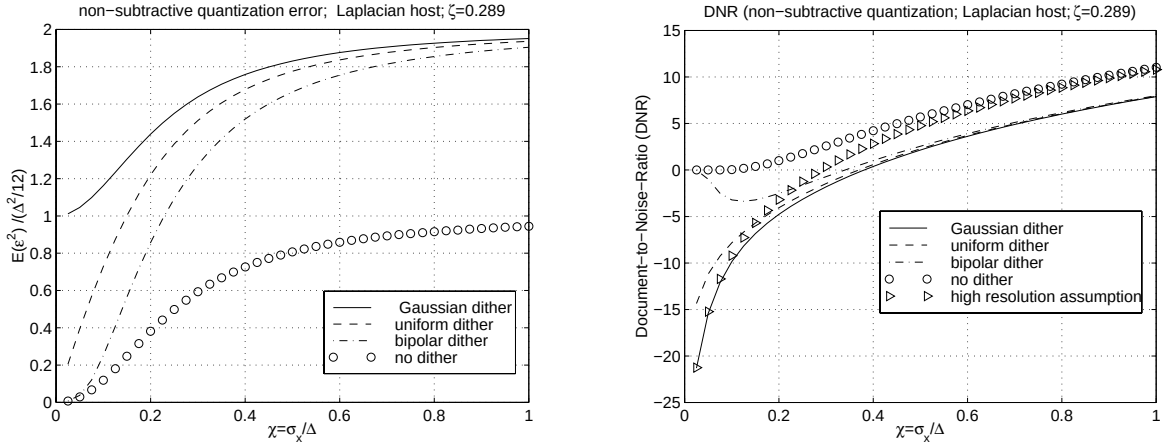

Figure 4. (Left) Normalized power of the non-subtractive DQE $\vec{\epsilon}$ for different dither distributions and $\zeta = \sigma_{\mathbf{d}}/\Delta = 1/\sqrt{12} = 0.289$; (right) document-to-noise ratio DNR $= 20\log_{10}\sigma_{\mathbf{x}}/\sigma_{\epsilon}$ after fingerprint embedding via dithered quantization.

3. CROSS-CORRELATION BETWEEN DIFFERENT QUANTIZATION ERRORS

In the previous section, dithered quantization was described and the corresponding mean squared (non-)subtractive DQE was analyzed. Thus, the embedding distortion of the proposed quantization watermarking scheme was characterized. In this section, we turn to the problem of distinguishing the differently quantized copies of the same original signal. We focus on the cross-correlation between the different quantization errors. In common additive watermarking schemes, it is easy to choose different watermarks independently. This is not the case for the proposed quantization watermarking scheme, since the cross-correlation

between different quantization errors may depend on the distribution of the dither signal and the host signal. We discuss theoretical results for fine dithered quantization, and some simulation results are presented for coarse quantization.

3.1. Fine quantization

In the case of fine quantization, the PDF of the host signal $\mathbf{x}$ is almost constant over the range of one quantization bin. Thus, the distribution of the host signal, e.g., Gaussian or Laplacian, has no influence on the statistics of the quantization noise. The expectation $\mathrm{E}\{\mathbf{xe}\}$ tends to zero, and, for dither signals drawn independently from the host signal, $\mathrm{E}\{\mathbf{x}\epsilon\}$ is almost zero, too. First of all, two uniform scalar quantizers with equal step size Δ, but constant offset $\alpha\Delta$, will be considered. An expression for the cross-correlation between their quantization errors $\mathbf{e_1}$ and $\mathbf{e_2}$ is derived. Then the cross-correlation between two (non-)subtractive DQEs achieved for two independent dither realizations d_1 and d_2 is computed by averaging over all possible effective quantizer shifts.

3.1.1. Two Quantizers with Constant Offset $\alpha\Delta$

We consider quantization with the prototype uniform scalar quantizer $\mathcal{Q} = \mathcal{Q}_1$ with step size Δ and the quantizer $\mathcal{Q}_2$ with step size Δ and offset $\alpha\Delta$, as shown on the left side of Fig. 5. α could take any real value, but, due to the symmetry of the uniform scalar quantizer, it is sufficient to consider $\alpha \in [0, 1)$. The subtractive DQE $e_2 = x - \mathcal{Q}_2[x]$ of the shifted quantizer $\mathcal{Q}_2$ can be expressed for every sample of the host signal x as a function of the subtractive DQE $e_1 = x - \mathcal{Q}_1[x]$ of the prototype quantizer $\mathcal{Q} = \mathcal{Q}_1$:

$$e_2 = \begin{cases} e_1 + \alpha\Delta & \text{for } e_1 \in \left[-\frac{\Delta}{2}, \frac{\Delta}{2} - \alpha\Delta\right] \\ e_1 + \alpha\Delta - \Delta & \text{for } e_1 \in \left[\frac{\Delta}{2} - \alpha\Delta, \frac{\Delta}{2}\right] \end{cases} \tag{11}$$

It is sufficient to investigate the quantization error for one quantization bin, since fine quantization is considered. The PDF of the subtractive DQE $\mathbf{e_1}$ can be approximated by $p_{\mathbf{e_1}}(e_1) = \mathrm{rect}(e_1/\Delta)$. For both quantizers, the mean squared error will be the same, specifically $\sigma^2_{\mathbf{e_1}} = \sigma^2_{\mathbf{e_2}} = \sigma^2_{\mathbf{e}} = \Delta^2/12$. Thus, the normalized cross-correlation between $\mathbf{e_1}$ and $\mathbf{e_2}$ is

$$\begin{aligned} \frac{\mathrm{E}\{\mathbf{e_1 e_2}\}}{\sigma^2_{\mathbf{e}}} &= \frac{1}{\sigma^2_{\mathbf{e}}\Delta} \int_{-\Delta/2}^{\Delta/2} e_1 \cdot e_2 \, \mathrm{d}e_1 = \frac{1}{\sigma^2_{\mathbf{e}}\Delta} \int_{-\Delta/2}^{\Delta/2-\alpha\Delta} e_1 \cdot (e_1 + \alpha\Delta) \, \mathrm{d}e_1 + \frac{1}{\sigma^2_{\mathbf{e}}\Delta} \int_{\Delta/2-\alpha\Delta}^{\Delta/2} e_1 \cdot (e_1 + \alpha\Delta - \Delta) \, \mathrm{d}e_1 \\ &= \frac{\Delta^2}{12\sigma^2_{\mathbf{e}}}(1 - 6\alpha + 6\alpha^2) = 1 - 6\alpha + 6\alpha^2 \end{aligned} \tag{12}$$

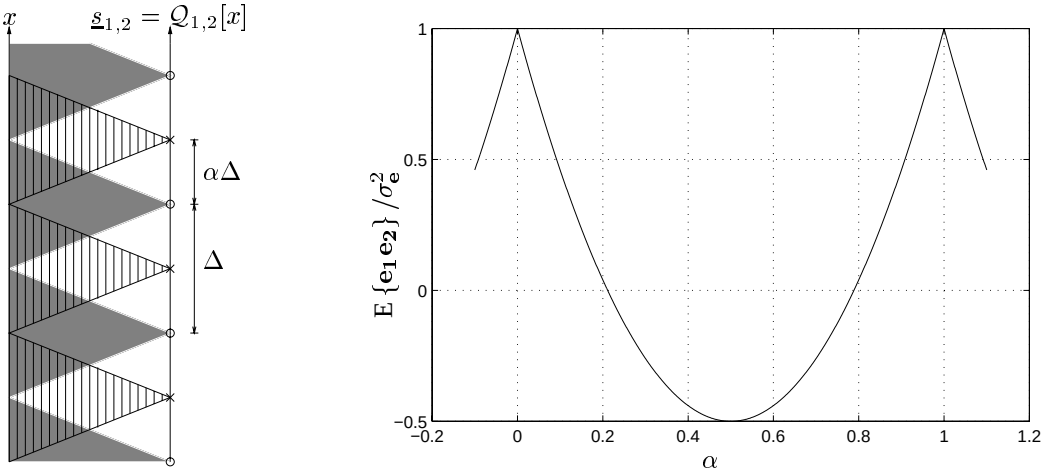

Figure 5. Cross-correlation between the quantization errors of two uniform scalar quantizers with constant offset $\alpha\Delta$

The quantization error cross-correlation as a function of α is shown in Fig. 5. The plot also indicates that (12) is periodic in α with the periodicity 1. Note that the cross-correlation result is identical if the offset $\alpha\Delta$ is not removed after quantization,

meaning $s_2 = \mathcal{Q}[x - \alpha\Delta]$ is considered instead of $\underline{s}_2 = \mathcal{Q}[x - \alpha\Delta] + \alpha\Delta$. Zero cross-correlation $\mathrm{E}\{\mathbf{e_1 e_2}\}$ is achieved for the quantizer offsets $\alpha_{0_{1,2}}\Delta$ with

$$\alpha_{0_{1,2}} = \frac{1}{2} \pm \sqrt{\frac{1}{12}} \approx 0.7887,\ 0.2113. \tag{13}$$

3.1.2. Quantization Error Cross-Correlation after Random Dithering

The cross-correlation between the DQE of two different (non-)subtractive dithered quantizers can be calculated by considering the effective offset between the representation levels of the subtractive dithered quantizers. With help of the PDF $p_\alpha(\alpha)$ of the effective offset, the cross-correlation can be computed by

$$\frac{\mathrm{E}\{\mathbf{e_1 e_2}\}}{\sigma_e^2} = \int_0^{0.5} (1 - 6\alpha + 6\alpha^2) p_\alpha(\alpha)\ \mathrm{d}\alpha, \tag{14}$$

where the periodicity and symmetry in α is exploited. Assuming independent dither signals $\vec{d}_1$ and $\vec{d}_2$ with the PDFs $p_{\mathbf{d_1}}(d_1)$ and $p_{\mathbf{d_2}}(d_2)$, the PDF of the effective offset difference can be computed by convolving the dither PDFs with each other, that is, $p_\alpha(\alpha) = p_{\mathbf{d_1}}(\alpha) \star p_{\mathbf{d_2}}(\alpha)$. Now, three example dither signals are investigated to illustrate this result. We assume in all examples that the PDFs $p_{\mathbf{d_1}}(d_1)$ and $p_{\mathbf{d_2}}(d_2)$ are identical.

Example 1: Bipolar dither with $d \in \{-\Delta/4, \Delta/4\}$

The effective offset is 0 in half of all cases and 0.5 otherwise. Hence, the cross-correlation between the DQE is $\mathrm{E}\{\mathbf{e_1 e_2}\}/\sigma_e^2 = 0.5 \cdot 1 + 0.5 \cdot (-0.5) = 0.25$.

Example 2: Continuous dither with uniform distribution in $[-\Delta/4, \Delta/4)$

Here, $p_\alpha(\alpha)$ is triangular with $p_\alpha(0) = 2$ and $p_\alpha(|\alpha| > 0.5) = 0$. Therefore, the cross-correlation between the DQE is $\mathrm{E}\{\mathbf{e_1 e_2}\}/\sigma_e^2 = 2\int_0^{0.5}(1 - 6\alpha + \alpha^2)(2 - 4\alpha)\mathrm{d}\alpha = 0.25$.

Example 3: Continuous dither with uniform distribution in $[-\Delta/2, \Delta/2)$

The effective offset difference will also have a triangular distribution, but with $p_\alpha(0) = 1$ and $p_\alpha(|\alpha| > 1) = 0$. However, due to the periodicity and symmetry in α, this PDF can be translated into a uniform distribution over all $\alpha \in [-\Delta/2, \Delta/2)$. Hence, the cross-correlation between the DQE is $\mathrm{E}\{\mathbf{e_1 e_2}\}/\sigma_e^2 = 2\int_0^{0.5}(1 - 6\alpha + \alpha^2)\mathrm{d}\alpha = 0$. This result agrees with the theory on dithered uniform scalar quantizers. As already mentioned in Section 2, Schuchman[10] showed that a uniform dither in $[-\Delta/2, \Delta/2)$ makes the subtractive DQE independent from the quantizer input.

The previous analysis shows that the cross-correlation between two (non-)subtractive DQE is strongly dependent on the statistics of the dither signal. We verified that zero cross-correlation between two quantization error signals, representing the fingerprints in our case, can be achieved for a continuous dither uniformly distributed in $[-\Delta/2, \Delta/2)$. Note that this is impossible for any random bipolar dither sequence, since for half of the samples the quantization error correlation is one and for the other half it cannot be lower than -0.5.

3.2. Coarse quantization

An analytic investigation of the cross-correlation between different DQEs is difficult for coarse quantization. This is due to the dependence of the DQE on the host signal and the dither distribution, which can be neglected only for fine quantization. Therefore, only simulation results are briefly discussed in this subsection. We chose a synthetic host signal with Laplacian distribution, applied dithered quantization with two different dither sequences, and computed the cross-correlation between both subtractive and non-subtractive DQEs. We investigate the case of two independently generated dither sequences $\vec{d}_1$ and $\vec{d}_2$, and the case of dependent dither sequences, specifically $\vec{d}_2 = -\vec{d}_1$. Again, Gaussian, uniform and bipolar dither distributions are considered. For fine quantization, the simulation results confirm the analytic results discussed in Section 3.1.

First of all, the case of two independent dither sequences is considered. We found that the results for $\mathrm{E}\{\mathbf{e_1 e_2}\}$ are almost identical to those of $\mathrm{E}\{\epsilon_1 \epsilon_2\}$, which corresponds to the same effect derived for fine quantization. The experiments showed that the normalized cross-correlation between the subtractive DQE increases rather than decreases for coarse dithered quantization. This effect is stronger for low-power dither signals, and agrees with intuition, which says that coarser quantization makes the quantized signals more similar rather than more different.

The case of two dependent dither sequences with $\vec{d}_2 = -\vec{d}_1$ is particularly important since such a signaling gives best watermark robustness when the dither is used for watermark detection. Obviously, it is difficult to design two dither sequences leading to $\epsilon_2 = -\epsilon_1$. Thus the relation between ϵ_1 and ϵ_2 is investigated for the case of dithering with $\vec{d}_2 = -\vec{d}_1$. Since $\mathrm{E}\{\epsilon_1\epsilon_2\} = \mathrm{E}\{\mathbf{e_1 e_2}\} + \mathrm{E}\{\mathbf{e_1 d_2}\} + \mathrm{E}\{\mathbf{e_2 d_1}\} + \mathrm{E}\{\mathbf{d_1 d_2}\}$, with $\mathrm{E}\{\mathbf{e_1 d_2}\} = \mathrm{E}\{\mathbf{e_2 d_1}\} \approx 0$ for sufficient fine quantization, the cross-correlation $\mathrm{E}\{\mathbf{e_1 e_2}\}$ determines also the cross-correlation $\mathrm{E}\{\epsilon_1\epsilon_2\}$. However, for coarse quantization the simulations show that the cross-correlations $\mathrm{E}\{\mathbf{e_1 d_2}\}$ and $\mathrm{E}\{\mathbf{e_2 d_1}\}$ increase and thus gives also an increased $\mathrm{E}\{\epsilon_1\epsilon_2\}$. This indicates again more similar DQEs for coarser quantization. The minimum measured $\mathrm{E}\{\epsilon_2\epsilon_1\}/\mathrm{E}\{{\epsilon_1}^2\}$ in all our experiments is -0.75, which is achieved for a bipolar dither distribution and fine quantization.

4. WATERMARK DETECTION AFTER COMBINED WATERMARKING AND COMPRESSION

4.1. General Detection Principle

Communication theory tells us that maximum a posteriori (MAP) detection forms the optimal decision rule when the costs for all possible errors are equal. When all codewords are sent with equal probability, MAP detection is equal to maximum likelihood (ML) detection. When Gaussian channel noise is assumed, a minimum-distance decoder can be derived from the ML rule. This step is critical in the case of watermark detection, since the channel noise (noise introduced by an attacker) might be non-Gaussian. Nevertheless, minimum-distance decoding usually works quite well, even for non-Gaussian channel noise, and thus is widely used. Finally, the minimum-distance detector can be translated into a correlation detector, under some weak assumption like equal power of all possibly transmitted signals. We will use a correlation detector, as in most watermarking schemes. However, some modifications are introduced to improve the detection performance where possible. As described in Section 1, the detector receives a signal $\vec{r} = \vec{x} + \vec{w} + \vec{v}$. The fingerprint detector has access to the original signal $\vec{x}$, and thus can subtract it from the received signal, which yields the pre-processed signal $\vec{y} = \vec{r} - \vec{x}$. For very strong attacks, a weighted subtraction of $\vec{x}$ would be better[12–14] since strong attacks also remove information about the original. However, for simplicity we neglect the weighting in this work.

Correlation detection means the decision is based on the value

$$c_k = \frac{1}{N}\sum_{n=1}^{N} y[n]w_k[n] = \frac{\vec{y}^t\vec{w}_k}{N} \quad \forall k \in \mathcal{K}. \tag{15}$$

Improved detection robustness can be gained by channel coding or the collusion-secure codes mentioned in Section 1, depending on the application. Here, $\mathcal{K}$ describes not the set of all possible fingerprints, but the set of possible indices for one code symbol. Only 2-ary signaling is considered, thus $\mathcal{K} = \{0, 1\}$. Further, an equiprobable distribution of k is assumed. Therefore, ML decision based on the measured correlation c_k forms the optimal detector, once we restrict ourself to use only c_k. In this case, the probability density functions (PDFs) of the hypotheses H_0 ($k = 0$ was sent) and H_1 ($k = 1$ was sent) describe the detection performance. Due to the summation over several samples in (15), the shape of the PDFs $p_c(c|H_0)$ and $p_c(c|H_1)$ can be approximated by a Gaussian distribution, which is completely determined by its mean and variance. Thus, the four parameters $\mu_{H_0} = \mathrm{E}\{c|H_1\}$, $\mu_{H_1} = \mathrm{E}\{c|H_1\}$, $\sigma_{H_0} = \mathrm{STD}\{c|H_0\}$ and $\sigma_{H_1} = \mathrm{STD}\{c|H_1\}$ are sufficient to design and describe the detection process.

4.2. Detection from Independent Channels

One important assumption when reducing ML decision to correlation detection is that the signal samples are I.I.D.. However, this is seldom true in practical watermarking schemes. For this reason it is useful to split the host signal into independent sub-signals, where each sub-signal is approximately I.I.D.. The watermark is embedded in each sub-signal separately. From the information-transmission viewpoint, the different sub-signals can be regarded as independent sub-channels. The detector computes the correlation (15) for each sub-channel, yielding c_i, and finally combines these values to make the decision between H_0 and H_1. We discussed this approach before[13] for the case of watermark detection after quantization attacks. In that work the decision between "watermark is present" and "watermark is not present" was considered and led to the assumption $\mu_{H_0,i} = 0,\ \forall i$. This assumption is not valid in our present case. Without derivation, we state here the extension for $\mu_{H_0,i} \neq 0$:

$$H_0: \quad \sum_{i=1}^{i_{\max}} c_i\omega_i > t \tag{16}$$

$$\text{with} \quad t = \sum_{i=1}^{i_{\max}} \frac{\mu_{H_0,i}^2 - \mu_{H_1,i}^2}{\sigma_i^2} \quad \text{and} \quad \omega_i = \frac{2(\mu_{H_0,i} - \mu_{H_1,i})}{\sigma_i^2}. \tag{17}$$

Here, $i_{\max}$ denotes the number of sub-channels, t is the decision threshold, and ω_i can be considered the weight for sub-channel i. The assumptions are the same as in our previous work, including that $\sigma^2_{H_0,i} = \sigma^2_{H_1,i} = \sigma^2_i$. When t and ω_i fit to the actual sub-channel characteristics, the bit error rate can be predicted to be

$$p_e = 0.5 \text{ erfc} \left(\sqrt{\sum_{i=1}^{i_{\max}} \frac{(\mu_{H_0,i} - \mu_{H_1,i})^2}{8\sigma_i^2}} \right), \tag{18}$$

where $\text{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty \exp(-\xi^2) \, \mathrm{d}\xi$. Note that the detection from sub-channels can be equivalent to the detection using whitening filters, as usually applied in communications. However, the sub-channel approach is more general since it contains also cases where, for instance, one sub-channel is the audio signal and the other the video signal of a movie. Using whitening filters in such a case seems to be awkward.

4.3. Exploiting Combined Fingerprinting and Compression at the Detector

In common additive watermarking schemes, $\vec{w}_k$ is the initially embedded noise sequence, which is drawn independently from the host signal $\vec{x}$. However, the additive embedding of such a sequence is incompatible with the desire for compressed documents, because the successive compression step already forms the first attack on the embedded watermark. We propose to define the non-subtractive DQE to be the watermark, that is $\vec{w}_{k,\mathtt{qwm}} = \vec{\epsilon}_k$. Note that the common additive scheme is achieved by defining the dither signal as the watermark $\vec{w}_{k,\mathtt{add}} = \vec{d}_k$. In the remainder of the article, $\vec{d}_2 = -\vec{d}_1$ is assumed. Our main results are similar for orthogonal dithering ($\vec{d}_2 \perp \vec{d}_1$), but omitted here due to space constraints. For the same reason, we further restrict the discussion to bipolar dither distributions with $d[n] \in \{\pm\frac{\Delta}{4}\}$.

As mentioned above, the fingerprints embedded by dithered quantization can be detected by computing the correlation (15), using either $\vec{w}_{k,\mathtt{qwm}}$ or $\vec{w}_{k,\mathtt{add}}$. However, in both cases we do not exploit the fact that many samples in both possible public signals $\vec{s}_1$ and $\vec{s}_2$ are identical. Let $\mathcal{N}_{\mathtt{same}}$ denote the subset of all sample indices $n \in \mathcal{N}$ with $s_1[n] = s_2[n]$. The complementary set is $\mathcal{N}_{\mathtt{diff}}$, thus $\mathcal{N} = \mathcal{N}_{\mathtt{same}} \cup \mathcal{N}_{\mathtt{diff}}$. All signal samples indexed by $\mathcal{N}_{\mathtt{same}}$ contain no information about the embedded fingerprint, so using them for the correlation measurement can only decrease the detection performance. This can be explained also by the sub-channel approach, discussed in Section 4.2. Let the samples indexed by $\mathcal{N}_{\mathtt{diff}}$ and those indexed by $\mathcal{N}_{\mathtt{same}}$ form two sub-channels. It is easy to see that $\omega_{\mathcal{N}_{\mathtt{same}}} = 0$, and thus the sub-channel for $\mathcal{N}_{\mathtt{same}}$ can be neglected completely. Note that the detector, knowing the host signal $\vec{x}$ and both possible subtractive DQE $\vec{\epsilon}_0$ and $\vec{\epsilon}_1$, also knows the sets $\mathcal{N}_{\mathtt{diff}}$ and $\mathcal{N}_{\mathtt{same}}$. Therefore, two additional detection methods are the correlation measurement over all samples indexed by $\mathcal{N}_{\mathtt{diff}}$, either using the non-subtractive DQE $\vec{w}_{k,\mathtt{qwm},\mathcal{N}_{\mathtt{diff}}}$ or the dither $\vec{w}_{k,\mathtt{add},\mathcal{N}_{\mathtt{diff}}}$. Hence, we consider detection using $\vec{w}_{k,\mathtt{qwm},\mathcal{N}}$, $\vec{w}_{k,\mathtt{add},\mathcal{N}}$, $\vec{w}_{k,\mathtt{qwm},\mathcal{N}_{\mathtt{diff}}}$ and $\vec{w}_{k,\mathtt{add},\mathcal{N}_{\mathtt{diff}}}$. The detection performance of all four proposed methods is compared in the next section.

5. DETECTION RESULTS FOR IMAGE FINGERPRINTS

We motivated quantization watermarking with the demand for fingerprinted and compressed images. In Section 2 and Section 3 the power of the quantization noise of dithered quantization and the cross-correlation between different dithered quantization errors was analyzed for synthetic signals. Now an example image fingerprinting scheme is considered, and the performance of the detection algorithms described in Section 4 is compared. The image fingerprinting scheme is not fully optimized. First of all, the image quality after dithered quantization has not been investigated in detail. However, the resuls from Section 2 indicate how dithered quantization performs compared to fixed quantization. Further, the image bit rate of the fingerprinted compressed image is not considered, which is an important factor for a practical scheme. On the other hand, we can assume that separate fingerprinting and compression cannot achieve lower image bit rates on average, since any embedded fingerprint can be considered as a kind of dither.

5.1. Description of the Experiment

We chose the popular JPEG compression scheme[15] for the example image fingerprinting and compression scheme. In JPEG compression, the image is transformed by an 8×8 block DCT, and the DCT coefficients are quantized by uniform quantizers. A table defines the quantizer step sizes for all 64 DCT coefficients per 8×8 block. The quantization table is parametrized by a quality factor $Q \in \{1, 2, \ldots, 100\}$, where highest visual quality and lowest compression is achieved for $Q = 100$. The quantizer step size is equal for all DCT coefficients for $Q = 100$ and $Q = 1$. In between, the quantizer step size may be different for different DCT coefficients; for instance, at high quality factors, only high-frequency coefficients are quantized coarsely.

We substitute the fixed quantization in the JPEG algorithm for some DCT coefficients by dithered quantization, where a bipolar dither with $d[n] \in \{\pm\frac{\Delta}{4}\}$ is used. A simple but useful heuristic for the selection of DCT coefficients to be fingerprinted is:

1. Apply JPEG compression to the original image with a quality factor $Q_{\texttt{emb}}$, where dithered quantization is used for all DCT coefficients.
2. Measure the resulting mean squared error $\text{MSE}_{\texttt{emb}}$ per DCT coefficient.
3. Apply JPEG compression to the original image with a quality factor $Q_{\min} < Q_{\texttt{emb}}$, where $Q_{\min}$ is the minimum quality factor achieving an acceptable image quality.
4. Measure the resulting mean squared error $\text{MSE}_{\min}$ per DCT coefficient.
5. Select those DCT coefficients for fingerprint embedding, where $\text{MSE}_{\min} > \text{MSE}_{\texttt{emb}}$ was measured.

Our presented results are obtained for $Q_{\texttt{emb}} = 80$ and $Q_{\min} = 69$, where the grayscale test image "Lenna" of size 256×256 was used. In this case 28, most low frequency DCT coefficients were selected. The experimental results presented below are obtained from 100 differently fingerprinted copies of Lenna. In all cases, 20 bits were embedded, where the actual watermark samples for one bit are spread over all DCT coefficients where dithered quantization is allowed.

So far, no attack has occured, since in our scheme the first compression step is combined with the fingerprint embedding. The robustness of the embedded fingerprints is analyzed for attacks by a second JPEG compression step, using quality factors from 10 to 100. Other possible attacks are simple additive white Gaussian noise (AWGN) attacks or the Gaussian test channel. Moulin and O'Sullivan[7] showed that the Gaussian test channel forms the optimal attack against watermarks in Gaussian host signals. However, the DCT coefficients have a distribution that is more peaked than a Laplacian. We found experimentally that in this case, quantization is a much stronger attack than the Gaussian test channel. Thus, we use JPEG compression as an attack, although we cannot show that it is the optimal attack (it is probably not). Watermark detection after quantization attacks has already been investigated.[12,13] The watermark robustness after JPEG compression is quite different for DCT coefficients of different frequency, such that a multi-channel approach, as described in Section 4.2, should be used, where each coefficient forms one sub-channel. Thus, up to 64 sub-channels can exist. However, not all of them may contain a watermark due to the constraint on the embedding distortion. For the given image size, 1024 DCT coefficients of each frequency exists. Thus, for such a sub-channel, each of the 20 watermark bits is embedded into roughly 50 coefficients. We consider only detection where the detector defined in (16) is optimized for the given attack. Although this might not possible in practical fingerprinting schemes, it gives us an upper bound on the performance of the considered detection methods.

5.2. Correlation Measurements

First of all, the distribution of the measured correlation values is investigated for one example. Since 20 watermark bits are embedded in 100 different simulations, 2000 correlation values are measured for each attack and each DCT-sub-channel. Fig. 6 depicts the measured PDFs $p_{\mathbf{c}}(c)$ for both hypotheses, where the attack was JPEG compression with quality $Q_{\texttt{attack}} = 50$. The presented correlations are measured for the 19th DCT coefficient in zigzag scan. All plots show that the assumption of a Gaussian PDF for $\mathbf{c}$ is at least roughly correct. The results shown in the upper plots of Fig. 6, were obtained when all watermark samples (indexed by $\mathcal{N}$) are used for the correlation. When detecting $\vec{w}_{\texttt{add},\mathcal{N}}$, the ratio $\mu_{H_0}/\mu_{H_1} = -1$ was measured and the standard deviations σ_{H_0} and σ_{H_1} are equal. An error rate of $p_e = 0.126$ was predicted. For the detection of $\vec{w}_{\texttt{qwm},\mathcal{N}}$, the ratio $\mu_{H_0}/\mu_{H_1} = -0.71$ was measured. The unequal means are also visible in the upper right plot of Fig. 6. Due to the differences in the variances of both hypotheses, different detection error rates $p_{e,H_0} = 0.063$ and $p_{e,H_1} = 0.077$ are predicted. This example shows that detecting $\vec{w}_{\texttt{qwm},\mathcal{N}}$ instead of $\vec{w}_{\texttt{add},\mathcal{N}}$ leads to better detection performance. On the other hand, the higher symmetry of the PDFs $p_{\mathbf{c}}(c)$ when detecting $\vec{w}_{\texttt{add},\mathcal{N}}$ might lead to more robust detection in practice. The main advantage of detecting $\vec{w}_{\texttt{add},\mathcal{N}}$ instead of $\vec{w}_{\texttt{qwm},\mathcal{N}}$ is that the threshold t can be set to zero, which simplifies a practical scheme.

The lower plots in Fig. 6 depict the correlation results when only selected watermark samples, namely those indexed by $\mathcal{N}_{\texttt{diff}}$, were used. We observe that for detection with $\vec{w}_{\texttt{qwm},\mathcal{N}_{\texttt{diff}}}$ both detection cases become more symmetric. Now, the ratio μ_{H_0}/μ_{H_1} is roughly -1, and the detection error rates decrease to $p_{e,H_0} = 0.053$ and $p_{e,H_1} = 0.059$. However, the improvement due to sample selection is much larger when detecting $\vec{w}_{\texttt{add},\mathcal{N}_{\texttt{diff}}}$. Here, the high symmetry between H_0 and H_1 remains, and the detection error rate decreases to $p_e = 0.045$. Thus, for the presented example, detection based on the dither samples $d[n]$ indexed by $\mathcal{N}_{\texttt{diff}}$ gives the lowest bit error rate. In the next subsection, we show that this result is consistent for other differently strong attacks and different DCT coefficients.

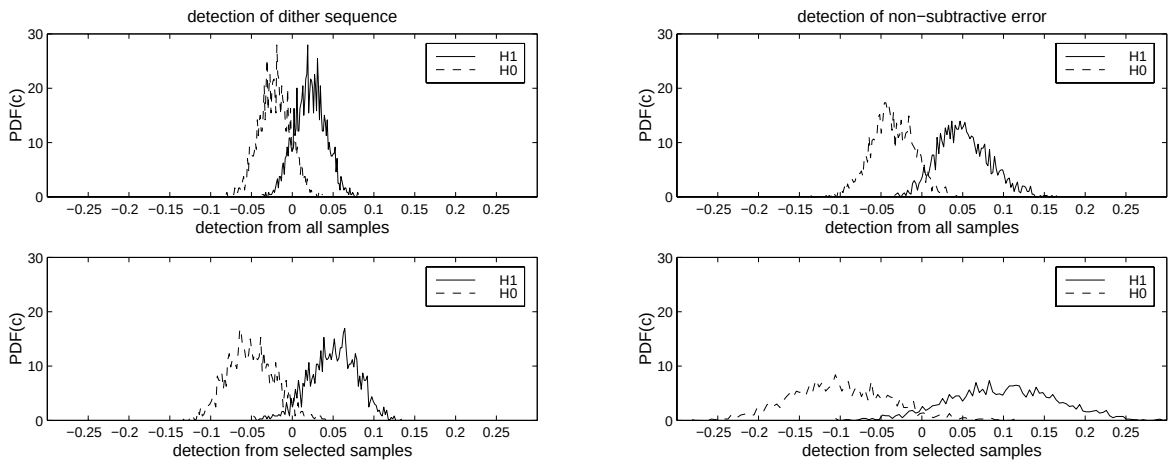

Figure 6. PDFs of the detected correlation value after an attack by coarse quantization; (left) detect dither; (right) detect non-subtractive DQE

5.3. Detection Error Probabilities

We compare the detection error probabilities for all considered correlation measurements for different attacks and different DCT coefficient. Fig. 7 depicts predicted and measured detection error probabilites, where a linear scale is used in the upper plots and a logarithmic scale is used in the lower plots. The left plots are for detection from the 5th DCT coefficient only, and the right plots show results obtained by combined detection from the DCT coefficients 5, 8 and 9. It is clearly visible in the upper plots that for strong attacks, detection with all dither samples ($\vec{w}_{\mathtt{add},\mathcal{N}}$) is less robust than the three other methods ($\vec{w}_{\mathtt{qwm},\mathcal{N}}$, $\vec{w}_{\mathtt{qwm},\mathcal{N}_{\mathtt{diff}}}$, $\vec{w}_{\mathtt{add},\mathcal{N}_{\mathtt{diff}}}$). Further, we observe that the measured and predicted error rates closely agree. It is also evident that adding information by combining the results from sub-channels 5, 8 and 9 increases the detection performance for all four correlation methods.

The lower plots, with the logarithmic scale, reveal the different performance of the four considered correlation methods. Detection with samples of $\vec{w}_{\mathtt{add}}$ in $\mathcal{N}_{\mathtt{diff}}$ gives the lowest bit error rate for all attacks. In addition, the detection performance when using three sub-channels instead of one improves considerably. The measured detection error rates fit not very accurately to the predicted values at low bit error rates. However, this effect occurs because of the limited number of experiments. The minimum measurable detection rate is $5 \cdot 10^{-4}$ for our experiments.

6. CONCLUSIONS

We propose to combine fingerprinting and compression by dithered quantization to achieve improved fingerprint detection. Therefore, uniform scalar dithered quantization is briefly reviewed in Section 2, and the effect of dithered quantization on the power of the quantization noise is analyzed theoretically. In Section 3, the cross-correlation between two different quantization errors is investigated analytically for fine quantization. For coarse quantization simulation results are discussed. In Section 4, we propose four methods for the fingerprint detection, which differ in the correlation measurement. The first two methods are correlation detection using the entire dither sequence or the entire non-subtractive DQE. In the other two methods the correlation with the dither sequence or the non-subtractive DQE is measured after selecting the signal samples for which the public signals $\vec{s}_0$ and $\vec{s}_1$ are not identical. The four detection methods are compared for an experimental image fingerprinting scheme. Correlating the entire received signal with the entire dither signal, which is equivalent to additive watermarking followed by quantization, performs worst. Instead, detecting the whole non-subtractive DQE gives significantly lower bit error rates. However, this scheme might lead to problems in practice, due to asymmetry of the measured correlation values. Finally, the best performance was achieved by correlating only the selected samples of the received signal with the corresponding dither samples. Thus, the knowledge about the applied dithered quantization can be exploited best by simply neglecting in the detection process all those signal samples where the public signals are identical due to quantization.

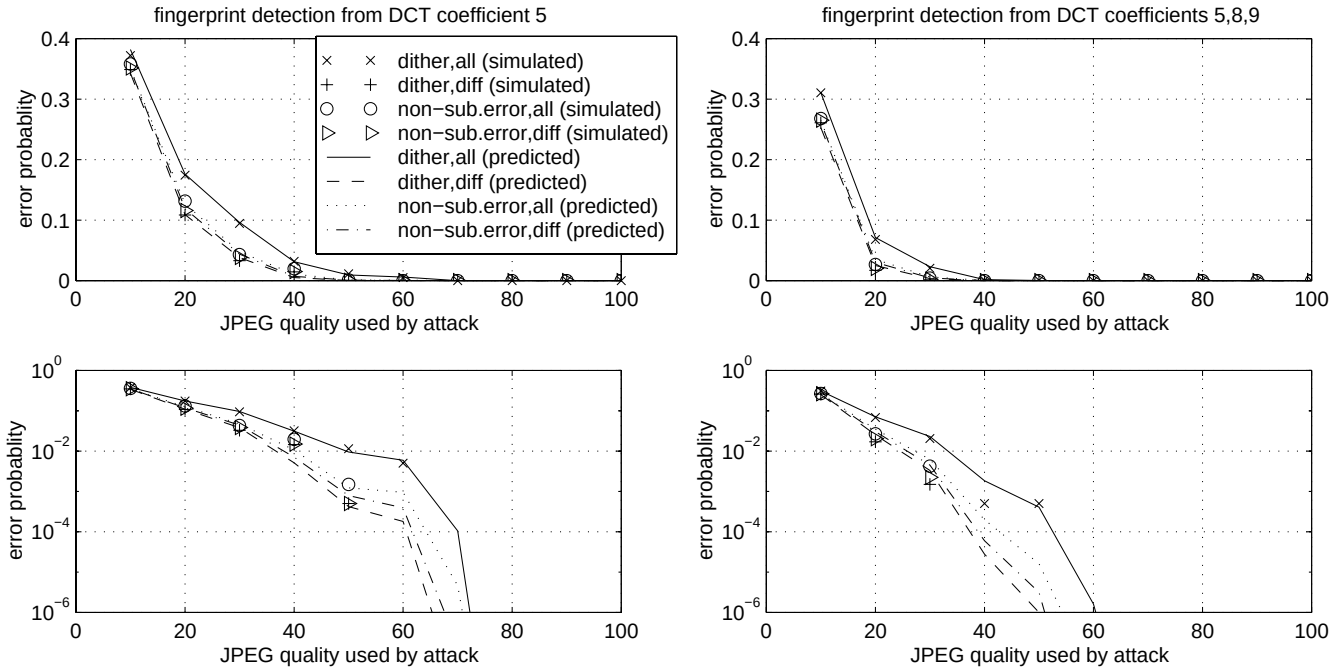

Figure 7. Simulated and predicted detection error rates; (left) detection from DCT coefficient 5; (right) detection from DCT coefficients 5, 8 and 9

REFERENCES

1. D. Boneh and J. Shaw, "Collusion secure fingerprinting for digital data," *IEEE Transactions on Information Theory* **44**, pp. 1897–1905, May 1998.
2. H.-J. Guth and B. Pfitzmann, "Error- and Collusion-Secure Fingerprinting for Digital Data," in *Workshop on Information Hiding*, (Dresden, Germany), October 1999.
3. X.-G. Xia, C. G. Boncelet, and G. R. Arce, "A multiresolution watermark for digital images," in *Proceedings of the IEEE International Conference on Image Processing 1997 (ICIP 97)*, vol. 1, pp. 548–551, (Santa Barbara, CA, USA), October 1997.
4. B. Chen and G. W. Wornell, "Dither modulation: a new approach to digital watermarking and information embedding," in *Proceedings of SPIE Vol. 3657: Security and Watermarking of Multimedia Contents*, (San Jose), January 1999.
5. M. Ramkumar, *Data Hiding in Multimedia: Theory and Applications*. PhD thesis, New Jersey Center for Multimedia Research, New Jersey Institute of Technology (NJIT), Kearny, NJ, USA, November 1999.
6. M. H. M. Costa, "Writing on Dirty Paper," *IEEE Transactions on Information Theory* **29**, pp. 439–441, May 1983.
7. P. Moulin and J. A. O'Sullivan, "Information-Theoretic Analysis of Information Hiding," September 1999. preprint.
8. R. M. Gray and T. G. Stockham, "Dithered quantizers," *IEEE Transactions on Information Theory* **39**, pp. 805–812, May 1993.
9. N. S. Jayant and P. Noll, *Digital Coding of Waveforms*, Prentice Hall, 1984.
10. L. Schuchman, "Dither signals and their effect on quantization noise," *IEEE Transaction on Communication Technology (COM)* **12**, pp. 162–165, December 1964.
11. R. Zamir and M. Feder, "On universal quantization by randomized uniform/lattice quantization," *IEEE Transactions on Information Theory* **38**, pp. 428–436, March 1992.
12. J. J. Eggers and B. Girod, "Quantization Effects on Digital Watermarks," *Signal Processing* , 1999. Submitted.
13. J. J. Eggers and B. Girod, "Watermark Detection after Quantization Attacks," in *Workshop on Information Hiding*, (Dresden, Germany), October 1999.
14. J. K. Su and B. Girod, "Fundamental permance limits of power-spectrum condition-compliant watermarks," in *Proceedings of SPIE Vol. 3971: Security and Watermarking of Multimedia Contents II*, (San Jose, Ca, USA), January 2000.
15. G. K. Wallace, "The JPEG still picture compression standard.," *Communications of the ACM* **34**, pp. 31–44, April 1991.