NOVEL TECHNIQUES FOR WATERMARK EXTRACTION IN THE DCT DOMAIN

J. R. Hernández , M. Amado and F. Pérez-González

Dept. Tecnologías de las Comunicaciones, ETSI Telecom., Universidad de Vigo, 36200 Vigo, Spain
email: jhernan@tsc.uvigo.es, fperez@tsc.uvigo.es

ABSTRACT

A novel DCT-domain watermark extraction procedure for still images that does not require the original image is presented. This method is based on the generalized Gaussian model, which includes as a special case the cross-correlation-based watermark detector structures, used so far in the literature. Optimal maximum likelihood (ML) structures are given, which allow to analytically assess the performance of watermarking methods in the DCT domain within a statistical framework. These original theoretical results are validated with experiments that show a considerable improvement over the existing watermark extraction techniques. The perceptual model used in the tests is also described.

1. INTRODUCTION

In recent years we have witnessed a striking proliferation of techniques for representation, storage and distribution of digital multimedia information. Unfortunately, these developments have also opened the gate to unathorized copying, distribution and manipulation of data, mostly images. Specialized and costly hardware may alleviate the problem of images duplication, at the price of a dramatic reduction in marketing possibilities –this is the cryptographic approach taken by pay TV channels, not foreseeable for scenarios such as Internet–. Watermarking techniques can at least ensure that ownership information is invisibly embedded into the image, thus preventing or deterring users from illegal uses.

Although many watermarking methods have sprouted over the few past years, even with commercial products available, the results up to date are quite discouraging, since there are freely available programs (e.g., unZign, Stirmark) that have succeeded in wiping the watermark away with little impact on the quality of the resulting image. Parallel to this, the lack of theoretical analyses in most of the available literature makes it difficult to know the actual limits in the performance of the various methods and to provide well-founded solutions which are the only way to eventually turn digital copyright protection into a mature discipline. In this paper we make a contribution in this direction by showing how watermarking in the DCT domain (the most commonly used) can be dramatically improved by carefully modeling the problem and designing the proper watermark detector. We will assume throughout the paper that the original image is not known. While knowledge of the original image greatly simplifies the extraction procedure [1] it also narrows the range of possible applications.

Let $x[\boldsymbol{n}]$ be a two-dimensional sequence representing the luminance of the original image, where $\boldsymbol{n} = (n_1, n_2)$. For the sake of readability, we will use in the sequel this vector notation to represent two-dimensional discrete indexes. Let $X[\boldsymbol{k}]$ be the result of applying a DCT transform to $x[\boldsymbol{n}]$ in a 8×8 pixels block basis. For copyright protection purposes, a watermark $W[\boldsymbol{k}]$ carrying some hidden information (owner and image identification number, transaction date, etc.) is added to the original image in the DCT domain, obtaining as a result the watermarked version $Y[\boldsymbol{k}] \triangleq X[\boldsymbol{k}] + W[\boldsymbol{k}]$.

In the watermarking technique we analyze in this paper, the watermark $W[\boldsymbol{k}]$ is generated in the DCT domain employing a 2-dimensional multipulse amplitude modulation scheme [2, 3]. In other words, $W[\boldsymbol{k}]$ can be expressed as the sum of N orthogonal pulses $\{P_i[\boldsymbol{k}]\}_{i=1}^N$

$$W[\boldsymbol{k}] = \sum_{i=1}^{N} b_i P_i[\boldsymbol{k}], \tag{1}$$

where the coefficients $\boldsymbol{b} \triangleq (b_1, \ldots, b_N)$ are used to encode the hidden message. The modulation pulses $\{P_i[\boldsymbol{k}]\}_{i=1}^N$ are generated as a function of a secret key K, only known by the copyright owner. They are expressed as

$$P_i[\boldsymbol{k}] = \begin{cases} \alpha[\boldsymbol{k}]s[\boldsymbol{k}], & \boldsymbol{k} \in \mathcal{S}_i \\ 0, & \text{otherwise}, \end{cases} \tag{2}$$

where $s[\boldsymbol{k}]$ is a key-dependent pseudorandom sequence such that $s[\boldsymbol{k}] \in \{-1, 1\}, \forall \boldsymbol{k}$, and the sets of indexes $\mathcal{T} \triangleq \{\mathcal{S}_i\}_{i=1}^N$ are also key-dependent and determine the spatial shape of the pulses. The sequence $\alpha[\boldsymbol{k}]$ is called the *perceptual mask* and indicates the maximum allowable magnitude of the alteration that the coefficient $X[\boldsymbol{k}]$ may suffer without achieving noticeable distortions. The sets $\{\mathcal{S}_i\}_{i=1}^N$ are assumed to be non-overlapping, i.e. $\mathcal{S}_i \cap \mathcal{S}_j = \emptyset, \forall i \neq j$, and sparsely spread over the whole image in a pseudorandom fashion to provide security and robustness against cropping [2, 3].

Given a watermarked image $Y[\boldsymbol{k}]$ and the secret key K, first the presence of a watermark for that key is tried to be detected in the so-called *watermark detection test*. If the result is positive, then the *watermark decoding* procedure obtains an estimate of the message $\boldsymbol{b}$. We will assume in this paper that no attacks aimed at desynchronizing the watermark are performed. However, both the synchronization and watermark detection problems can be tackled within the statistical framework presented in sections 2 to 4. The perceptual model used in our particular watermarking scheme is given in Section 5. Section 6 is devoted to experimental results, while Section 7 presents our conclusions and future lines of research.

2. STATISTICAL MODEL

Detector structures usually proposed for hidden information decoding in DCT-domain spread spectrum data hiding techniques are based on the crosscorrelation between the watermarked image

Work partially funded by CICYT under contract TIC-96-0500-C10-10

$Y[\boldsymbol{k}]$ and the pseudorandom sequence $s[\boldsymbol{k}]$. This scheme would be appropriate if noise –in watermarking, the original image– followed a Gaussian distribution. However, the Gaussian assumption is inaccurate for DCT coefficients of common images. Some authors have proposed the generalized Gaussian probability density function (pdf)

$$f_x(x) = A\,\mathrm{e}^{-|\beta\, x|^c}. \tag{3}$$

as an alternative leading to improved statistical models [4]. Note that the Gaussian and the Laplacian pdf's are just special cases of this expression, given by $c = 2$ and $c = 1$, respectively. Previous works in this field show that DCT coefficients at low frequencies are reasonably well-modeled by a generalized Gaussian distribution with $c = 1/2$. Coefficients at high frequencies are better approximated by a Gaussian distribution and sometimes by a Laplacian distribution.

The parameters A and β in Eq. (3) can be expressed as

$$\beta = \frac{1}{\sigma}\left(\frac{\Gamma(3/c)}{\Gamma(1/c)}\right)^{1/2};\ A = \frac{\beta\, c}{2\Gamma(1/c)}, \tag{4}$$

where σ is the standard deviation. Hence, the pdf is completely specified by c and σ. Let us define the sequence

$$C_{i,j}[k_1,k_2] \triangleq X[8k_1+i, 8k_2+j],\ \ i,j \in \{0,\ldots,7\},$$

which results if we take the (i,j)-th DCT coefficient of every block. We will model each of these 64 sequences as the output of a two-dimensional i.i.d. random process whose marginal distribution follows Eq. (3), with parameters $c(i,j)$ and $\sigma(i,j)$. Let us also define the sequences $c[\boldsymbol{k}]$ as

$$c[\boldsymbol{k}] \quad \triangleq \quad c(k_1\ mod\ 8,\ k_2\ mod\ 8)$$

and $\sigma[\boldsymbol{k}]$ in a similar fashion. Thus, these two sequences indicate the parameters c and σ associated with each sample $X[\boldsymbol{k}]$.

3. WATERMARK DECODER

Let us assume that M possible different messages can be encoded with the vector $\boldsymbol{b} = (b_1,\cdots,b_N)$ and let $\boldsymbol{b}_l, l \in \{1,\cdots,M\}$ denote the codeword associated to one of those messages. Also, let $W_l[\boldsymbol{k}], l \in \{1,\ldots,M\}$ be the watermark obtained from $\boldsymbol{b}_l = (b_{l,1},\ldots,b_{l,N})$ using Eq. (1). Then, assuming the i.i.d. generalized Gaussian model for $X[\boldsymbol{k}]$, it can be easily shown that the optimum decoder in the ML sense is the one that chooses the index $l \in \{1,\ldots,M\}$ verifying

$$\sum_{\boldsymbol{k}} \frac{|Y[\boldsymbol{k}]-W_m[\boldsymbol{k}]|^{c[\boldsymbol{k}]} - |Y[\boldsymbol{k}]-W_l[\boldsymbol{k}]|^{c[\boldsymbol{k}]}}{\sigma[\boldsymbol{k}]^{c[\boldsymbol{k}]}} > 0,\ \ \forall\, m \neq l.$$

Assuming that $b_{l,i} \in \{-1,1\}, \forall l \in \{1,\ldots,M\}, i \in \{1,\ldots,N\}$, this optimization problem is equivalent to finding the codeword $\boldsymbol{b}_l$ which maximizes the expression $\sum_{i=1}^{N} b_{l,i}\, r_i$, where the coefficients r_i are sufficient statistics for the detection problem and are defined as

$$r_i \triangleq \sum_{\boldsymbol{k}\in\mathcal{S}_i} \frac{|Y[\boldsymbol{k}]+\alpha[\boldsymbol{k}]s[\boldsymbol{k}]|^{c[\boldsymbol{k}]} - |Y[\boldsymbol{k}]-\alpha[\boldsymbol{k}]s[\boldsymbol{k}]|^{c[\boldsymbol{k}]}}{\sigma[\boldsymbol{k}]^{c[\boldsymbol{k}]}}.$$

When a binary antipodal constellation is used to encode $M = 2^N$ possible messages, the ML detector structure is equivalent to a bit-by-bit hard decisor, so the outputs of the decoder are

$$\hat{b}_i = \mathrm{sgn}(r_i),\ \ i \in \{1,\ldots,N\}.$$

Now let us analyze the performance of the watermark decoding process in terms of the *probability of bit error* P_b. Obviously, performance results strongly depend on image characteristics, so we will obtain P_b conditioned to a given original image $X[\boldsymbol{k}]$ or, in other words, the probability of getting a bit error when a secret key is taken at random and is applied in both the watermarking and decoding processes. In this context, $X[\boldsymbol{k}]$ will be regarded as a deterministic signal while the sequence $s[\boldsymbol{k}]$ and the sets $\mathcal{T} = \{\mathcal{S}_i\}_{i=1}^{N}$ will be modeled statistically.

If the pseudorandom sequence $s[\boldsymbol{n}]$ is modeled as an i.i.d. two-dimensional random process with marginal pdf $f_s(s)$, then, each sufficient statistic r_i is the sum of $|\mathcal{S}_i|$ statistically independent contributions ($|\mathcal{S}_i|$ is the cardinality of the set $\{\boldsymbol{k}, P_i[\boldsymbol{k}] \neq 0\}$). Hence, by central limit theorem arguments, $\boldsymbol{r} \triangleq (r_1,\ldots,r_N)$ can be accurately approximated as the output of a vector Gaussian channel. Therefore, the probability of error conditioned to $X[\boldsymbol{k}]$ can be expressed as a function of the first and second order moments of $r_1,\ldots,r_N$. Let us define the two-dimensional sequence

$$r[\boldsymbol{k}] \triangleq |Y[\boldsymbol{k}]+\alpha[\boldsymbol{k}]s[\boldsymbol{k}]|^{c[\boldsymbol{k}]} - |Y[\boldsymbol{k}]-\alpha[\boldsymbol{k}]s[\boldsymbol{k}]|^{c[\boldsymbol{k}]},$$

extracted from Eq. (3).

If the tiling generation process is such that each index $\boldsymbol{k} \in \mathbb{N}^2$ belongs to $\mathcal{S}_i$ with probability $1/N$ for all $i \in \{1,\ldots,N\}$ and assignments of indices to sets are performed independently, i.e. $Pr\{\boldsymbol{k} \in \mathcal{S}_i, \boldsymbol{m} \in \mathcal{S}_j\} = Pr\{\boldsymbol{k} \in \mathcal{S}_i\}Pr\{\boldsymbol{m} \in \mathcal{S}_j\}, \forall \boldsymbol{k} \neq \boldsymbol{m}, i,j \in \{1,\ldots,N\}$, then after some algebraic manipulations it can be proven [10] that

$$E[r_i] = \frac{1}{N}\sum_{\boldsymbol{k}} \frac{E[r[\boldsymbol{k}]]}{\sigma[\boldsymbol{k}]^{c[\boldsymbol{k}]}}. \tag{5}$$

$$Var(r_i) = \frac{1}{N}\sum_{\boldsymbol{k}} \frac{Var(r[\boldsymbol{k}])}{\sigma[\boldsymbol{k}]^{2c[\boldsymbol{k}]}} + \frac{N-1}{N^2}\sum_{\boldsymbol{k}} \frac{E^2[r[\boldsymbol{k}]]}{\sigma[\boldsymbol{k}]^{2c[\boldsymbol{k}]}}. \tag{6}$$

Assume that $b_i = 1$. Then, $Y[\boldsymbol{k}] = X[\boldsymbol{k}] + \alpha[\boldsymbol{k}]s[\boldsymbol{k}], \forall \boldsymbol{k} \in \mathcal{S}_i$, and, as a consequence,

$$r[\boldsymbol{k}] = |X[\boldsymbol{k}]+2\alpha[\boldsymbol{k}]s[\boldsymbol{k}]|^{c[\boldsymbol{k}]} - |X[\boldsymbol{k}]|^{c[\boldsymbol{k}]}.$$

If $s[\boldsymbol{k}]$ follows a discrete uniform two-level distribution, $s[\boldsymbol{k}] \in \{-1,1\}$, it can be easily shown that the mean and variance of $r[\boldsymbol{k}]$ are

$$\begin{aligned}
E[r[\boldsymbol{k}]] \quad &= \quad \frac{1}{2}\Big[\Big(|X[\boldsymbol{k}]|+2\alpha[\boldsymbol{k}]\Big)^{c[\boldsymbol{k}]} \\
&\quad + \Big||X[\boldsymbol{k}]|-2\alpha[\boldsymbol{k}]\Big|^{c[\boldsymbol{k}]}\Big] - |X[\boldsymbol{k}]|^{c[\boldsymbol{k}]} \\
Var(r[\boldsymbol{k}]) \quad &= \quad \frac{1}{4}\Big[\Big(|X[\boldsymbol{k}]|+2\alpha[\boldsymbol{k}]\Big)^{c[\boldsymbol{k}]} \\
&\quad - \Big||X[\boldsymbol{k}]|-2\alpha[\boldsymbol{k}]\Big|^{c[\boldsymbol{k}]}\Big]^2.
\end{aligned}$$

These expressions can be applied in equations (5) and (6) to compute the moments of r_i. When $b_i = -1$, it can be verified that

$Var(r_i)$ is given by Eq. (6) and $E[r_i]$ is negative and its absolute value is given by Eq. (5). When a binary antipodal constelation with $M = 2^N$ is used to encode the hidden message, the probability of bit error P_b of the ML decoder (a bit-by-bit hard decisor) is $P_b = Q\,(SNR)$, where $Q(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_x^{\infty} \mathrm{e}^{-t^2/2} \mathrm{d}t$ and the signal to noise ratio SNR is defined as

$$SNR \triangleq \frac{E[r_i]}{\sqrt{Var(r_i)}}. \tag{7}$$

4. WATERMARK DETECTOR

Now let us analyze the watermark detection test, in which we have to decide whether a given image contains a watermark generated with a certain key. The watermark detection problem can be mathematically formulated as the binary hypothesis test

$$\begin{array}{ll} H_1: & Y[\boldsymbol{k}] = X[\boldsymbol{k}] + W[\boldsymbol{k}] \\ H_0: & Y[\boldsymbol{k}] = X[\boldsymbol{k}] \end{array} \tag{8}$$

where $X[\boldsymbol{k}]$ is the original image, not available during the test, and $W[\boldsymbol{k}]$ is a watermark generated from the secret key K that is tested. If the watermark carries hidden information, it is not the goal of the watermark detection test to estimate the hidden message; this task is left to the decoding process. Therefore, we must take into account the uncertainty about the value of the codeword vector $\boldsymbol{b}$ when designing the detector. The optimum ML (*Maximum Likelihood*) decision rule for the test formulated above is

$$\Lambda(Y) \underset{H_0}{\overset{H_1}{\gtrless}} \eta, \tag{9}$$

where η is the decision threshold and $\Lambda(Y)$ is the likelihood function

$$\Lambda(Y) = \frac{1}{M} \sum_{l=1}^{M} \frac{f(Y \mid H_1, \boldsymbol{b}_l)}{f(Y \mid H_0)} \tag{10}$$

If we assume that the coefficients of the original image $X[\boldsymbol{k}]$ follow the generalized Gaussian model studied in Sect. 2, and that the watermark does not carry hidden information, in other words, that there is only one pulse ($N = 1$) and it is modulated by a known value $b_1 = 1$, then the log-likelihood function $l(Y) \triangleq \ln \Lambda(Y)$ has the form

$$l(Y) = \sum_{\boldsymbol{k}} \beta[\boldsymbol{k}]^{c[\boldsymbol{k}]} \left(|Y[\boldsymbol{k}]|^{c[\boldsymbol{k}]} - |Y[\boldsymbol{k}] - \alpha[\boldsymbol{k}] s[\boldsymbol{k}]|^{c[\boldsymbol{k}]} \right). \tag{11}$$

where $\beta[\boldsymbol{k}]$ is the parameter β in the generalized Gaussian pdf for the coefficient $X[\boldsymbol{k}]$ (it can be obtained from $c[\boldsymbol{k}]$ and $\sigma[\boldsymbol{k}]$ using Eq. (4)).

Let us now analyze the performance of the watermark detection test conditioned to a certain original image. For this purpose, we will characterize statistically $l(Y)$ for each of the two hypothesis when we assume that $s[\boldsymbol{k}]$ is the only random element in the watermarking system. When H_0 is true, we have that $Y[\boldsymbol{k}] = X[\boldsymbol{k}],\ \forall \boldsymbol{k}$. Therefore,

$$l(Y) = \sum_{\boldsymbol{k}} \beta[\boldsymbol{k}]^{c[\boldsymbol{k}]} \left(|X[\boldsymbol{k}]|^{c[\boldsymbol{k}]} - |X[\boldsymbol{k}] - \alpha[\boldsymbol{k}] s[\boldsymbol{k}]|^{c[\boldsymbol{k}]} \right),$$

which is a sum of statistically independent terms. Hence, applying the central limit theorem we can infer that $l(Y)$ is approximately Gaussian. Assuming that $s[\boldsymbol{k}]$ is an i.i.d. two-dimensional random sequence with a discrete marginal distribution with two equiprobable levels, $s[\boldsymbol{k}] \in \{-1, 1\},\ \forall \boldsymbol{k}$, then we can easily prove that the mean and variance of $l(Y)$ conditioned to H_0 are [10]

$$\begin{aligned} E[l(Y) \mid H_0] \;=\; & \sum_{\boldsymbol{k}} \beta[\boldsymbol{k}]^{c[\boldsymbol{k}]} |X[\boldsymbol{k}]|^{c[\boldsymbol{k}]} \\ & - \frac{1}{2} \sum_{\boldsymbol{k}} \beta[\boldsymbol{k}]^{c[\boldsymbol{k}]} \Big(|X[\boldsymbol{k}] + \alpha[\boldsymbol{k}]|^{c[\boldsymbol{k}]} \\ & + |X[\boldsymbol{k}] - \alpha[\boldsymbol{k}]|^{c[\boldsymbol{k}]} \Big) \end{aligned} \tag{12}$$

$$\begin{aligned} Var(l(Y) \mid H_0) \;=\; & \frac{1}{4} \sum_{\boldsymbol{k}} \beta[\boldsymbol{k}]^{2c[\boldsymbol{k}]} \Big(|X[\boldsymbol{k}] + \alpha[\boldsymbol{k}]|^{c[\boldsymbol{k}]} \\ & - |X[\boldsymbol{k}] - \alpha[\boldsymbol{k}]|^{c[\boldsymbol{k}]} \Big)^2. \end{aligned} \tag{13}$$

Similarly, we can prove that $l(Y)$ conditioned to H_1 is approximately Gaussian with mean and variance

$$E[l(Y) \mid H_1] = -E[l(Y) \mid H_0] \tag{14}$$

$$Var(l(Y) \mid H_1) = Var(l(Y) \mid H_0) \tag{15}$$

Let us define $m_1 \triangleq E[l(Y) \mid H_1]$ and $\sigma_1^2 \triangleq Var(l(Y) \mid H_1)$. If H_1 is decided in the detection test when $l(Y) > \eta$, then the probabilities of false alarm (P_F) and detection (P_D) are

$$P_F = Q\left(\frac{\eta + m_1}{\sigma_1}\right), \qquad P_D = Q\left(\frac{\eta - m_1}{\sigma_1}\right). \tag{16}$$

Let us define the following "signal to noise ratio"

$$SNR_1 \triangleq \frac{m_1^2}{\sigma_1^2}. \tag{17}$$

If we denote by $Q^{-1}(P_F)$ the value $x \in \mathbb{R}$ such that $Q(x) = P_F$, then it can be easily proved, by examining the expressions in (16), that

$$P_D = Q\left(Q^{-1}(P_F) - 2\sqrt{SNR_1}\right). \tag{18}$$

Hence, the ROC (*Receiver Operating Characteristic*) of the watermark detector depends exclusively on the value of SNR_1. Obviously, the larger the value of SNR_1, the larger the P_D associated with a certain P_F and the better, as a consequence, the performance of the detector.

5. PERCEPTUAL MODEL

In Eqs. (1,2) the watermark $W[\boldsymbol{k}]$ depends on a perceptual mask $\alpha[\boldsymbol{k}]$ that multiplies the pseudorandom sequence $s[\boldsymbol{k}]$. This perceptual mask determines the maximum amplitude distortion that each coefficient of the original image may suffer while satisfying the invisibility constraint. A good psychovisual model in the DCT-domain (with 8x8 blocks) is capital to render the sequence $\alpha[\boldsymbol{k}]$. For our work we have followed the model proposed in [5, 6] that has been also applied to derive adaptive quantization matrices for the JPEG algorithm [7]. This model has been here simplified by disregarding the so-called *contrast-masking effect* for which the perceptual mask at a certain coefficient depends on the amplitude of the coefficient itself. Consideration of this effect constitutes a future line of research. On the other hand, the *background intensity effect*, for which the mask depends on the magnitude of the DC coefficient (i.e., the background), has been taken into account.

The so-called *visibility threshold* $T(i,j)$, $i \in \{0,\cdots,7\}$, $j \in \{0,\cdots,7\}$, determines the maximum allowable magnitude of an invisible alteration of the (i,j)-th DCT coefficient and can be approximated in logarithmic units by the following quadratic function with parameter K

$$\log T(i,j) = \log\left(\frac{T_{min}(f_{i,0}^2+f_{0,j}^2)^2}{(f_{i,0}^2+f_{0,j}^2)^2-4(1-r)f_{i,0}^2f_{0,j}^2}\right) + K\left(\log\sqrt{f_{i,0}^2+f_{0,j}^2}-\log f_{min}\right)^2,$$

where $f_{i,0}$ and $f_{0,j}$ are respectively the vertical and horizontal spatial frequencies (in cycles/degree) of the DCT-basis functions, T_{min} is the minimum value of $T(i,j)$, associated to the spatial frequency f_{min}, and r is taken as 0.7 following [5]. The threshold $T(i,j)$ can be corrected for each block by considering the DC coefficient $X_{0,0}$ and the average luminance of the screen $\overline{X}_{0,0}$ (1024 for an 8-bit image) in the following way

$$T'(i,j) = T(i,j)\left(\frac{X_{0,0}}{\overline{X}_{0,0}}\right)^{a_T}.$$

Note that the actual dependence of $X_{0,0}$ on the block indices has been dropped in the notation for conciseness. Following [5], the parameters used in our scheme have been set to $a_T = 0.649$, $f_{min} = 3.68$ cycles/degree, $T_{min} = 1.1548$ and $K = 1.728$. Once the corrected threshold value $T'_{i,j}$ has been obtained, the perceptual mask is calculated as

$$\alpha[k_1,k_2] = 4\left(1+\frac{\delta(l_1)}{\sqrt{2}+1}\right)\left(1+\frac{\delta(l_2)}{\sqrt{2}+1}\right)\gamma\cdot T'(l_1,l_2) \quad (19)$$

where $l_1 = k_1 \ mod \ 8$, $l_2 = k_2 \ mod \ 8$ and $\gamma < 1$ is a scaling factor that allows to introduce a certain degree of conservativeness in the watermark due to those effects that have been overlooked (e.g., spatial masking in the frequency domain [8]). The remaining factors in (19) allow to express the corrected threshold in terms of DCT coefficients instead of luminances.

6. EXPERIMENTAL RESULTS

In order to validate the theoretical analysis presented in previous sections, we have watermarked the well-known image *Lena* (256 x 256 pixels) following the method described in Sect. 1, modifying only 22 coefficients in the mid-frequency range (low frequency coefficients have very low capacity, i.e., slight modifications become quite visible; high frequency coefficients can be easily erased by compression algorithms).

To analyze the performance of the *watermark decoding* process, we watermarked the image *Lena* with 100 different keys for different bit rates, measured in terms of the number of coefficients altered by each information bit, and computed the resulting bit error rate (BER). Figure 1 shows one of the watermarks used in the experiment. In Figure 2 both empirical and theoretical results for different values of the generalized Gaussian parameter c are plotted. Note that the parameter γ in Eq. (19) has been set to 1/5 –so the watermark is well below the visibility level– in order to produce statistically significant results. The actual performance is substantially better, but the qualitative conclusions remain the same. As can be inferred from Figure 2 and also from Figure 3, where the SNR in Eq. (7) is plotted for different values of c, good results are obtained in the range $1/2 \leq c \leq 1$. Interestingly enough, the performance for $c = 2$, corresponding to the cross-correlation-based detector used so far in the literature [9], suffers a severe deterioration, corresponding to a drop of more than 6 dB in the SNR (cf. Fig. 3).

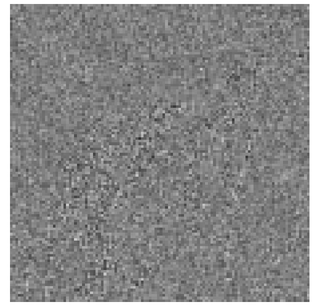

Figure 1: *One of the watermarks used in the tests.*

	$c = 1/2$	*Laplace*	*Gaussian*
Empirical	29.38	29.07	21.39
Theoretical	29.34	28.71	20.78

Table 1: Empirical and theoretical signal to noise ratio SNR_1 (in dB) in the watermark detection test.

Although not directly discussed here, our analysis can be somewhat straightforwardly extended to the case of JPEG compression. Figure 4 shows the theoretical BER obtained when image ‘Lena’ is watermarked (with a 100 bits hidden message) and later compressed with JPEG to a percentage of its original quality. As it can be seen, in this case, the Laplacian detector ($c = 1$) performs slightly better than the one with $c = 1/2$. The Gaussian (cross-correlation) detector, not shown in the figure, leads to a much higher BER. The curves labeled as ‘Optimum’ correspond to a detector specifically designed for the JPEG compression attack.

To measure the performance of the *watermark detection* test, we have watermarked the image *Lena* with 1000 different keys using “pure” watermarks carrying no additional information. In Table 1 we show both empirical and theoretical values of the “signal to noise ratio” SNR_1 for the image *Lena* and detector structures based on three values of c. As we have already discussed, SNR_1 completely determines the shape of the ROC. We can see that the theoretical approximations accurately fit the empirical data. Empirical measures of P_F and P_D, not shown here, clearly validate the accurateness of approximations in Eq. (16) [10]. Besides this, it is clear that substantial gains in performance are obtained by abandonning the Gaussian statistical assumption.

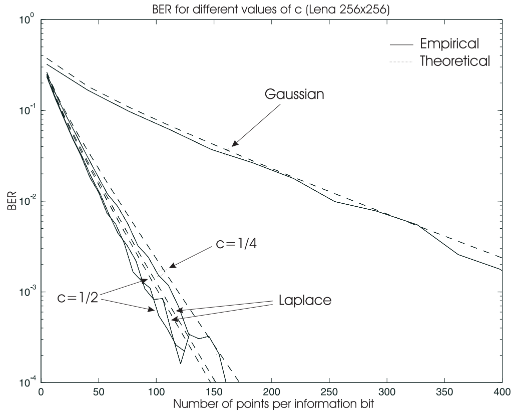

Figure 2: *BER as a function of the pulse size for Lena (256×256).*

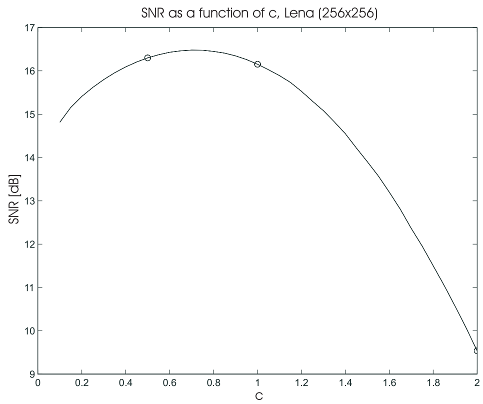

Figure 3: *SNR as a function of c for Lena (256×256).*

7. CONCLUSIONS AND FURTHER WORK

Novel structures based on the use of generalized Gaussian models have been proposed for the ML detection of DCT-domain watermarks embedded in still images. By considering these models, we have been able to dramatically improve the performance of the cross-correlation-based detectors that have been used up to date. In any case, we also have presented a theoretical analysis that allows to assess the performance of DCT-based methods, measured in terms of the bit error rate and the probabilities of false alarm and detection, for a given image. The Gaussian detector is simply a particular case of the generalized model, so the analytical results given here are directly applicable. One immediate extension of our analysis is the consideration of channel codes which have been already shown to considerably improve on spatial-domain watermarking methods [11]. Another future line of research consists in fitting a generalized Gaussian model (from a discrete set of parameters c and σ) to the DCT coefficients histogram so as to decide upon and apply the optimal detector structure.

8. REFERENCES

[1] I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Transactions on Image Processing*, vol. 6, pp. 1673–1687, December 1997.

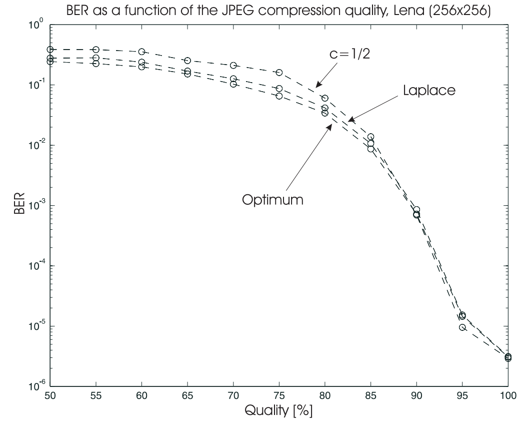

Figure 4: *BER as a function of JPEG final quality for Lena.*

[2] J. R. Hernández, F. Pérez-González, J. M. Rodríguez, and G. Nieto, "Performance analysis of a 2d-multipulse amplitude modulation scheme for data hiding and watermarking of still images," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 510–524, May 1998.

[3] J. R. Hernández, F. Pérez-González, and J. M. Rodríguez, "The impact of channel coding on the performance of spatial watermarking for copyright protection," in *Proc. ICASSP'98*, vol. 5, (Seattle, Washington (USA)), pp. 2973–2976, May 1998.

[4] R. J. Clarke, *Transform Coding of Images*. Academic Press, 1985.

[5] A. J. Ahumada and H. A. Peterson, "Luminance-model-based DCT quantization for color image compression," in *Human Vision, Visual Processing, and Digital Display III (Proc. of the SPIE)* (B. E. Rogowitz, ed.), 1992.

[6] J. A. Solomon, A. B. Watson, and A. J. Ahumada, "Visibility of DCT basis functions: Effects of contrast masking," in *Proc. Data Compression Conference*, (Snowbird, Utah (USA)), pp. 361–370, IEEE Computer Society Press, 1994.

[7] A. B. Watson, "Visual optimization of DCT quantization matrices for individual images," in *Proc., AIAA Computing in Aerospace 9*, (San Diego, California (USA)), pp. 286–291, American Institute of Aeronautics and Astronautics, 1993.

[8] A. N. Netravali and B. G. Haskell, *Digital Pictures. Representation, Compression and Standards*. New York: Plenum Press, 1995.

[9] M. Barni, F. Bartolini, V. Cappellini, and A. Piva, "A DCT-domain system for robust image watermarking," *Signal Procesing*, vol. 66, pp. 357–372, May 1998.

[10] J. R. Hernández, *Técnicas de Sellado Invisible para la Protección de los Derechos de Autor en Imágenes Digitales*. PhD thesis, Universidade de Vigo, 1998.

[11] J. R. Hernández, J. M. Rodríguez, and F. Pérez-González, "Improving the performance of spatial watermarking of images using channel coding." Accepted for publication in Signal Procesing, Elsevier.