DCT-Domain Watermarking Techniques for Still Images: Detector Performance Analysis and a New Structure

Juan R. Hernández, *Associate Member, IEEE*, Martín Amado, and Fernando Pérez-González, *Member, IEEE*

Abstract—In this paper, a spread-spectrum-like discrete cosine transform domain (DCT domain) watermarking technique for copyright protection of still digital images is analyzed. The DCT is applied in blocks of 8 × 8 pixels as in the JPEG algorithm. The watermark can encode information to track illegal misuses. For flexibility purposes, the original image is not necessary during the ownership verification process, so it must be modeled by noise. Two tests are involved in the ownership verification stage: watermark decoding, in which the message carried by the watermark is extracted, and watermark detection, which decides whether a given image contains a watermark generated with a certain key. We apply generalized Gaussian distributions to statistically model the DCT coefficients of the original image and show how the resulting detector structures lead to considerable improvements in performance with respect to the correlation receiver, which has been widely considered in the literature and makes use of the Gaussian noise assumption. As a result of our work, analytical expressions for performance measures such as the probability of error in watermark decoding and probabilities of false alarm and detection in watermark detection are derived and contrasted with experimental results.

Index Terms—Copyright protection, discrete cosine transform, signal detection, spread spectrum communication.

I. INTRODUCTION

THE DEVELOPMENT achieved during the last decades by both image processing techniques and telecommunication networks have facilitated the current explosion of applications that considerably simplify the acquisition, representation, storage, and distribution of images in digital format. Since protocol and network design is oriented to digital data delivery, most contents providers are rapidly transforming their archives to a digital format. However, all these advances have also made it possible to produce digital copies identical to the original with the greatest ease. In addition, unauthorized manipulation or reuse of information has become so common that this emerging creative process has been put in danger.

Although cryptography is an effective tool against the illegal digital distribution problem, it has to be coupled with specialized hardware in order to avoid direct access to data in digital format, something that it is not only costly but would also re-

duce the marketing possibilities for the service provider. It is in this scenario where watermarking techniques can prove to be useful. A digital watermark is a distinguishing piece of information that is adhered to the data that it is intended to protect. For obvious reasons, we will only consider here invisible watermarks that protect rights by embedding ownership information into the digital image in an unnoticeable way. This imperceptibility constraint is attained by taking into account the properties of the human visual system (HVS), which in turn helps to make the watermark more robust to most types of attacks. In fact, robustness of the watermark is a capital issue, since it should be resilient to standard manipulations, both intentional or unintentional and should also withstand multiple watermarking to facilitate the tracking of the subsequent transactions an image is subject to. Ideally, watermark destruction attacks should affect the original image in a similar way. However, to what extent the creator regards an image as his/hers is a difficult question whose answer depends on the application.

Watermarking, like cryptography, needs secret keys to identify legal owners; furthermore, most applications demand extra information to be hidden in the original image (steganography). This information may consist in ownership identifiers, transaction dates, serial numbers, etc., that play a key role when illegal providers are being tracked. Closely related to the embedment of this information is the extraction (*watermark decoding*) process whenever in possession of the secret key. In most cases of interest, there will be a certain probability of error for the hidden information which can be used as a measure of the performance of the system. Clearly, this probability will increase with the number of information bits in the message.

Another important problem is the so-called *watermark detection* problem, that consists in testing whether the image was watermarked by a certain key owner. Thus, the watermark detection problem can be formulated as a binary hypothesis test, for which a probability of false alarm (i.e., of deciding that a given key was used while it was not) and a probability of detection (i.e., of correctly deciding that a given key was used) can be defined. Note that the existence of "false positives" would dramatically reduce the credibility of the system, so the probability of false alarm should be kept to a very small value by adequately designing the test so it produces acceptable probabilities of detection.

Existing algorithms for watermarking still images usually work either in the spatial domain [1]–[8] or in a transformed domain [7], [9]–[15].

In this paper, we will deal with the analysis of watermarking methods in the discrete cosine transform (DCT) domain, which
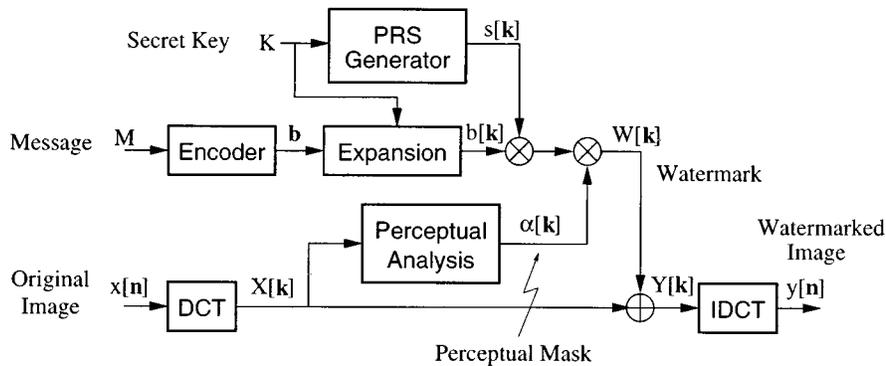
Fig. 1.   Block diagram of the watermark embedding process.

is widely used in compression applications and consequently in digital distribution networks. Moreover, invisibility constraints are easier to impose when working in such domain. We will assume that the original image is not present during the watermark detection and information decoding processes, since this would be hard to manage when huge quantities of images need be compared by intelligent agents searching the net for unauthorized copies.

The main goal of this paper is to present a novel analytical framework that allow to assess the performance of a given watermarking method in the DCT-domain. New results have been achieved by resorting to a theoretical model of the problem that permits to derive optimal detector and decoding structures. With those results, it is possible to determine how much information one can hide in a given image for a certain probability of error or what is the probability of correctly deciding that a given image has been watermarked. Moreover, these results are not tied to a specific key but rather they are averaged over the set of possible keys. One relevant by-product of the rigorous analysis presented in this paper has been the discovery of new detector and decoder structures that outperform the existing ones, usually based in the calculation of the correlation coefficient (sometimes called *similarity*) between the image and the watermark. As shown in this paper, these correlation structures, that have been somewhat taken for granted in the previous literature in DCT-domain watermarking, would be optimal only if the DCT coefficients followed a Gaussian distribution. However, as many authors have pointed out, midfrequency DCT coefficients can be quite accurately modeled by a so-called generalized Gaussian distribution. This distribution is the basis for the development of the structures that will be proposed. Recent works in image watermarking have produced a vast amount of algorithms; unfortunately, in most cases no theoretical results assessing their performance are known. In this regard, the theoretical approach presented here is crucial if we want to compare different methods and know what are their fundamental limits of performance [16].

This paper is organized as follows. In Sections II and III, the watermark generation and the watermark verification processes, respectively, are presented and formulated in mathematical notation. In Section IV, the perceptual model used in our experiments to guarantee that the watermarks generated are invisible is briefly discussed. The generalized Gaussian model for the sta-

tistical representation of images in the DCT domain is reviewed in Section V. Then, in Sections VI and VII, the watermark decoding (extraction) and detection problems, respectively, are analyzed and optimum detector structures and expressions for performance measures are obtained. Theoretical results are contrasted against empirical data obtained through experimentation in Sections VI-D and VII-B for the watermark decoder and detector, respectively.

II. WATERMARK EMBEDDING PROCESS

Let $x[\boldsymbol{n}] = x[n_1, n_2]$, $0 \leq n_1 < N_1$, $0 \leq n_2 < N_2$ be a two-dimensional (2-D) discrete sequence representing the luminance component of a sampled image with size $N_1 \times N_2$ pixels. In the sequel we will always use vector notation (in boldface typesetting) to represent 2-D indexes. The watermark is generated as a DCT-domain signal $W[\boldsymbol{k}]$ employing a technique similar to the direct-sequence spread spectrum modulation schemes used in communications (Fig. 1). In this paper, we will assume that the DCT is applied in blocks of 8×8 pixels, as in the JPEG algorithm [17]. We will allow the watermark to carry a *hidden message* M with information that could be used, for instance, to identify the intended recipient of the protected image. This message is mapped by an encoder into an N-dimensional codeword vector $\boldsymbol{b} = (b_1, \cdots, b_N)$. Then, a 2-D sequence $b[\boldsymbol{k}]$ is generated in what we call an expansion process, by repeating each element b_i, $i \in \{1, \cdots, N\}$ of the codeword in a different set $\mathcal{S}_i$ of points in the DCT-domain discrete grid, in such a way that the whole transformed image $X[\boldsymbol{k}]$ is covered.

For security purposes, it is possible to introduce uncertainty about the DCT coefficients altered by each codeword element by introducing an interleaving stage, consisting in a key-dependent pseudorandom permutation of the samples of $b[\boldsymbol{k}]$. In the sequel we will denote by $\mathcal{S}_i$ the set of DCT coefficients associated with the codeword coefficient b_i after the interleaving stage.

The resulting signal is multiplied in a pixel-by-pixel manner by the output $s[\boldsymbol{k}]$ of a pseudorandom sequence (PRS) generator with an initial state which depends on the value of the secret key. Finally, the spread spectrum signal is further multiplied by what we call a perceptual mask $\alpha[\boldsymbol{k}]$, which is basically used to amplify or attenuate the watermark at each DCT coefficient so that the watermark energy is maximized while the alterations suffered by the image are kept invisible. The perceptual mask $\alpha[\boldsymbol{k}]$

is obtained through a perceptual analysis of the original image $X[\boldsymbol{k}]$, based on a perceptual model in which frequency masking properties of the HVS are taken into account. The perceptual model used in the experiments presented in this paper is discussed in more detail in Section IV. The resulting signal $W[\boldsymbol{k}]$ is the watermark and is added to the original image $X[\boldsymbol{k}]$ to obtain the watermarked version $Y[\boldsymbol{k}] = X[\boldsymbol{k}] + W[\boldsymbol{k}]$.

III. WATERMARK VERIFICATION PROCESS

Once a watermarked image is distributed, the rights holder should be able to verify the copyright information to prove his authorship and possibly trace illegal misuses. In Fig. 2 we have represented in block diagram form the different steps involved in the watermark verification process, which will be analyzed throughout this paper.

First, given an image $z[\boldsymbol{n}]$, the 8×8 block-wise DCT transform $Z[\boldsymbol{k}]$ is computed. The watermark verification process is comprised of two tests. First, a *watermark detector* decides whether $Z[\boldsymbol{k}]$ contains a watermark generated from the secret key K. If it does contain such a watermark, then the *watermark decoder* obtains an estimate of the message M (see Section II). For each of these tests (Sections VII and VI, respectively), a set of sufficient statistics is computed from $Z[\boldsymbol{k}]$. Since we assume that the original image is not available during watermark verification, we must treat it as additive noise. In fact, we will use statistical models of DCT coefficients of common images (Section V) to analytically derive the appropriate sufficient statistics (Sections VI-A and VII-A). Suitable values for the parameters of these models must be either fixed previously or adaptively estimated from $Z[\boldsymbol{k}]$ (Section VI-C). As we will see, in order to be able to compute the sufficient statistics, it is necessary to know the pseudorandom sequence $s[\boldsymbol{k}]$ and the perceptual mask $\alpha[\boldsymbol{k}]$ used in the watermark embedding process (see Section II). Note that it is impossible to know exactly the perceptual mask since the original image is not accessible. However, a good estimate can be obtained from $Z[\boldsymbol{k}]$ using exactly the same perceptual analysis as in the watermark embedding unit, given the low perceptual distortion that the watermark introduces.

Once the sufficient statistics are computed, a decision is made in the watermark detector after evaluating the log-likelihood function for the binary hypothesis test (Section VII-A) and comparing the result with a threshold η. The sufficient statistics for watermark decoding are passed to a decoder, which obtains an estimate of the message M denoted by $\hat{M}$.

IV. PERCEPTUAL ANALYSIS

As we have seen in the previous section, during the watermark generation process a perceptual mask $\alpha[\boldsymbol{k}]$ multiplies the pseudorandom sequence to guarantee that the alterations introduced by the watermark are invisible. We will assume that the perceptual mask $\alpha[\boldsymbol{k}]$ indicates the maximum admissible amplitude of the alteration suffered by the DCT coefficient $X[\boldsymbol{k}]$. To obtain the mask $\alpha[\boldsymbol{k}]$, it is necessary to use a psychovisual model in the DCT domain. For the work presented in this paper, we have followed the model proposed by Ahumada *et al.* [18], [19], which has been applied by Watson to the adaptive compu-

tation of quantization matrices in the JPEG algorithm [20]. 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* $T(i, j)$, $i, j \in \{0, \cdots, 7\}$, 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 :

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

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 [18]. This mathematical model is not valid for the DC coefficient, so both i and j cannot be simultaneously zero. 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-b image) in the following way:

$$T'(i, j) = T(i, j) \left(\frac{X_{0,0}}{\overline{X}_{0,0}} \right)^{a_T}. \quad (2)$$

Note that the actual dependence of $X_{0,0}$ on the block indices has been dropped in the notation for conciseness. Following [18], 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

$$\alpha[k_1, k_2] = 4 \cdot \left(1 + \left(\sqrt{2} - 1 \right) \delta(l_1) \right) \cdot \left(1 + \left(\sqrt{2} - 1 \right) \delta(l_2) \right) \cdot \gamma \cdot T'(l_1, l_2) \quad (3)$$

where $l_1 = k_1 \bmod 8$, $l_2 = k_2 \bmod 8$, $\delta(.)$ is the Kronecker function, 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 [21]). The remaining factors in (3) allow to express the corrected threshold in terms of DCT coefficients instead of luminances (see [18]).

In all the experiments shown in this paper the perceptual mask obtained by following the procedure explained above has been divided by five. The only purpose of this modification is to artificially introduce a degradation of the performance so that we can obtain empirical estimates of performance measures in a reasonable amount of simulation time, within a range of values of system parameters (e.g., the pulse size) that allows us to compare results for different images. Hence, in a practical situation we should expect considerable better performance.

V. STATISTICAL MODELS OF THE DCT COEFFICIENTS

In this section, we briefly describe a statistical model that has been proposed to better characterize the DCT coefficients of common images. As we will see in the following sections, the use of these models when designing the watermark detector and decoder will lead to considerable improvements in performance [16] with respect to the correlating detector structure, usually proposed in previous literature, which would be optimum only if the original image (i.e., the channel additive noise) could be modeled as a Gaussian stochastic process.

Let $X[\boldsymbol{k}]$ be the 2–D sequence that we obtain after applying the DCT to the original image $x[\boldsymbol{n}]$, in blocks of 8×8 pixels. Let $C_{i,j}[k_1, k_2]$ be the 2–D sequence that results when we extract the DCT coefficient (i, j) of each block, i.e.,

$$C_{i,j}[k_1, k_2] \triangleq X[8k_1 + i, 8k_2 + j], \qquad i, j \in \{0, \cdots, 7\}. \tag{4}$$

A detailed discussion on the shapes of the histograms for different types of images can be found in [22]. Here we will only review the results in [22] that are relevant for our work. Considering the irregularity of its histogram and its highly image-dependent nature, the dc coefficient cannot be approximated accurately by any closed-form pdf (probability density functions). The ac coefficients, as shown in [22] and [23], can be reasonably well approximated by zero-mean *generalized Gaussian* pdf's, given by the expression

$$f_x(x) = Ae^{-|\beta x|^c}. \tag{5}$$

Both A and β can be expressed as a function of c and the standard deviation σ

$$\beta = \frac{1}{\sigma} \left(\frac{\Gamma(3/c)}{\Gamma(1/c)} \right)^{1/2}, \quad A = \frac{\beta c}{2\Gamma(1/c)}. \tag{6}$$

Therefore, this distribution is completely specified by two parameters, c and σ. Note that the Gaussian and Laplacian distributions are just special cases of this generalized pdf, given by $c = 2$ and $c = 1$, respectively.

One of the most important conclusions in [22] is that the DCT coefficients at low frequencies cannot be well approximated by a Gaussian pdf, not even by a Laplace pdf. For instance, in [22] it is shown how for a given test image, with characteristics similar to the well-known Lena image, the coefficients at low frequencies can be reasonably well modeled by a generalized Gaussian pdf with $c = 1/2$. This fact has important consequences, as we will see later, in the derivation of detector structures involved in watermark verification procedures.

For the sake of generality in the theoretical derivations, we will assume that the parameters c and σ can take different values for each coefficient (i, j) of the blockwise DCT. In other words, we will model each sequence $C_{i,j}[\boldsymbol{k}]$ as the output of an independent identically distributed (i.i.d.) generalized Gaussian stochastic process with parameters $c(i, j)$ and $\sigma(i, j)$. We will also define, in order to simplify the notation, the 2–D sequences representing the parameters that characterize the pdf of the DCT coefficients, as

$$c[\boldsymbol{k}] \triangleq c(k_1 \bmod 8, \quad k_2 \bmod 8) \tag{7}$$

$$\sigma[\boldsymbol{k}] \triangleq \sigma(k_1 \bmod 8, \quad k_2 \bmod 8). \tag{8}$$

VI. HIDDEN INFORMATION DECODING

We have just seen that, even though the DCT coefficients (excluding the DC coefficient) have histograms that can be reasonably well approximated by generalized Gaussian distributions, the pure Gaussian model is in many cases far from being the most adequate representation. This fact raises the question of whether the detector structure based on a correlation receiver, usually proposed in the previous literature on DCT-domain watermarking of images [24], is appropriate for either watermark decoding or watermark detection. In the following sections we will obtain the optimum *maximum likelihood* (ML) decoder structures which result when a generalized Gaussian noise model is assumed and will also analyze performance for different values of c, including the Gaussian case ($c = 2$). We will see that in some cases, such as the pure Gaussian distribution, improvements can be attained just by disregarding DCT coefficients with high amplitudes. We will also show how to estimate the parameters of the generalized Gaussian distribution that are necessary for a practical implementation of the newly proposed structures.

A. Optimum Decoder for the Generalized Gaussian Model

Let us assume that the vector $\boldsymbol{b} = (b_1, \cdots, b_N)$ can be used to encode one out of L possible messages ($\log_2 L$ bits) and that the set of vectors $\{\boldsymbol{b}_l\}_{l=1}^{L}$ indicates which codeword corresponds to each message. During the watermark verification process, the hidden information decoder obtains an estimate of the hidden message M embedded in the watermarked image $Y[\boldsymbol{k}]$. If we assume that the messages are equiprobable, the detector which minimizes the average probability of error is given by the ML test, in which the estimated message is the one satisfying

$$\ln \frac{f(Y|\boldsymbol{b}_l)}{f(Y|\boldsymbol{b}_m)} > 0, \qquad \forall m \neq l. \tag{9}$$

Considering the statistical model that has been assumed for the sequences $C_{i,j}[\boldsymbol{k}]$, and assuming also that these sequences are statistically independent (this assumption is justified given the uncorrelating properties of the DCT for common images), the ML test is equivalent to seeking the index $l \in \{1, \cdots, L\}$ which satisfies

$$\sum_{\boldsymbol{k}} \frac{|Y[\boldsymbol{k}] - W_m[\boldsymbol{k}]|^{c[\boldsymbol{k}]} - |Y[\boldsymbol{k}] - W_l[\boldsymbol{k}]|^{c[\boldsymbol{k}]}}{\sigma[\boldsymbol{k}]^{c[\boldsymbol{k}]}} > 0,$$
$$\forall m \neq l \tag{10}$$

where $W_m[\boldsymbol{k}]$ and $W_l[\boldsymbol{k}]$ are the watermarks generated from the codewords $\boldsymbol{b}_m$ and $\boldsymbol{b}_l$ following the mechanism presented in Section II. We can easily prove that the coefficients

$$r_i \triangleq \sum_{\boldsymbol{k} \in \mathcal{S}_i} \frac{|Y[\boldsymbol{k}] + \alpha[\boldsymbol{k}]s[\boldsymbol{k}]|^{c[\boldsymbol{k}]} - |Y[\boldsymbol{k}] - \alpha[\boldsymbol{k}]s[\boldsymbol{k}]|^{c[\boldsymbol{k}]}}{\sigma[\boldsymbol{k}]^{c[\boldsymbol{k}]}} \tag{11}$$

are sufficient statistics for the ML hidden information decoding process, which is equivalent to seeking the codeword

$\boldsymbol{b}_l = (b_{l,1}, \cdots, b_{l,N}), l \in \{1, \cdots, L\}$ that maximizes the expression

$$\sum_{i=1}^{N} b_{l,i}\, r_i. \tag{12}$$

For the case that a binary antipodal constellation with $L = 2^N$ codewords such that $b_{m,i} \in \{-1, 1\}, \forall m \in \{1, \cdots, L\}, i \in \{1, \cdots, N\}$ is used, this test is equivalent to a bit-by-bit hard decoder

$$\hat{b}_i = \operatorname{sgn}(r_i), \qquad \forall i \in \{1, \cdots, N\}. \tag{13}$$

It is interesting to analyze the performance of the hidden information decoder for a given original image. Specifically, we will derive the bit error rate, measured as the proportion of secret keys for which an error occurs while decoding a certain bit. In order to do so, first it is necessary to statistically characterize the coefficients $\{r_i\}_{i=1}^{N}$ assuming a fixed original image. In Appendix A we show that these coefficients can be modeled as the output of an *additive white Gaussian noise* (AWGN) channel, $r_i = ab_i + n_i, i \in \{1, \cdots, N\}$, where n_i are i.i.d. samples of zero-mean white Gaussian noise with variance σ_n^2. The parameters of this equivalent vector channel are

$$a = E[r_i] = \frac{1}{N} \sum_{\boldsymbol{k}} \frac{E[r[\boldsymbol{k}]]}{\sigma[\boldsymbol{k}]^{c[\boldsymbol{k}]}} \tag{14}$$

$$\sigma_n^2 = \operatorname{Var}[r_i] = \frac{1}{N} \sum_{\boldsymbol{k}} \frac{\operatorname{Var}(r[\boldsymbol{k}])}{\sigma[\boldsymbol{k}]^{2c[\boldsymbol{k}]}} + \frac{N-1}{N^2} \sum_{\boldsymbol{k}} \frac{E^2[r[\boldsymbol{k}]]}{\sigma[\boldsymbol{k}]^{2c[\boldsymbol{k}]}} \tag{15}$$

where

$$E[r[\boldsymbol{k}]] = \tfrac{1}{2} \left[(|X[\boldsymbol{k}]| + 2\alpha[\boldsymbol{k}])^{c[\boldsymbol{k}]} + |\,|X[\boldsymbol{k}]| - 2\alpha[\boldsymbol{k}]|^{c[\boldsymbol{k}]} \right] - |X[\boldsymbol{k}]|^{c[\boldsymbol{k}]} \tag{16}$$

$$\operatorname{Var}(r[\boldsymbol{k}]) = \tfrac{1}{4} \left[(|X[\boldsymbol{k}]| + 2\alpha[\boldsymbol{k}])^{c[\boldsymbol{k}]} - |\,|X[\boldsymbol{k}]| - 2\alpha[\boldsymbol{k}]|^{c[\boldsymbol{k}]} \right]^2. \tag{17}$$

Therefore, when codewords form a binary antipodal constellation, the probability of bit error for the bit-by-bit hard decoder is

$$P_b = Q\left(\sqrt{\operatorname{SNR}}\right) \tag{18}$$

where $\operatorname{SNR} \triangleq a^2/\sigma_n^2$ and $Q(\cdot)$ is defined as $Q(x) \triangleq (1/\sqrt{2\pi}) \int_x^{\infty} e^{-t^2/2}\, dt$.

B. Decoder with Point Elimination

If we examine typical histograms of DCT coefficients, we can observe that there are some samples in the tails with relatively high amplitudes that the generalized Gaussian model cannot model adequately. This deviation with respect to the theoretical model in the tails of the histogram leads us to think that including these samples in the optimum detector for that model may result in a loss of performance in the watermark decoder.

To explore this possibility, next we will define a new statistic following the same expression as that derived in Section VI-A for a generalized Gaussian noise model, but in which only those DCT coefficients $Y[\boldsymbol{k}]$ whose amplitude is below a certain threshold are taken into account. We will express this threshold as the product of a constant K_e and the perceptual mask $\alpha[\boldsymbol{k}]$. The resulting statistic is

$$r_i = \sum_{\boldsymbol{k} \in \mathcal{A} \cap \mathcal{S}_i} \frac{|Y[\boldsymbol{k}] + \alpha[\boldsymbol{k}]s[\boldsymbol{k}]|^{c[\boldsymbol{k}]} - |Y[\boldsymbol{k}] - \alpha[\boldsymbol{k}]s[\boldsymbol{k}]|^{c[\boldsymbol{k}]}}{\sigma[\boldsymbol{k}]^{c[\boldsymbol{k}]}} \tag{19}$$

where $\mathcal{A} \triangleq \{\boldsymbol{k}: |Y[\boldsymbol{k}]| \leq K_e\alpha[\boldsymbol{k}]\}$. Following similar arguments as in Appendix A it can be shown that the statistics $r_i, i \in \{1, \cdots, N\}$ can be modeled as the output of an AWGN channel, whose parameters are now

$$a = \frac{1}{N} \sum_{\boldsymbol{k} \in \mathcal{B}_1} \frac{E[r[\boldsymbol{k}]]}{\sigma[\boldsymbol{k}]^{c[\boldsymbol{k}]}} + \frac{1}{N} \sum_{\boldsymbol{k} \in \mathcal{B}_2} \frac{E[r'[\boldsymbol{k}]]}{\sigma[\boldsymbol{k}]^{c[\boldsymbol{k}]}} \tag{20}$$

$$\sigma_n^2 = \frac{1}{N} \sum_{\boldsymbol{k} \in \mathcal{B}_1} \frac{\operatorname{Var}(r[\boldsymbol{k}])}{\sigma[\boldsymbol{k}]^{2c[\boldsymbol{k}]}} + \frac{N-1}{N^2} \sum_{\boldsymbol{k} \in \mathcal{B}_1} \frac{E^2[r[\boldsymbol{k}]]}{\sigma[\boldsymbol{k}]^{2c[\boldsymbol{k}]}}$$
$$+ \frac{1}{N} \sum_{\boldsymbol{k} \in \mathcal{B}_2} \frac{\operatorname{Var}(r'[\boldsymbol{k}])}{\sigma[\boldsymbol{k}]^{2c[\boldsymbol{k}]}} + \frac{N-1}{N^2} \sum_{\boldsymbol{k} \in \mathcal{B}_2} \frac{E^2[r'[\boldsymbol{k}]]}{\sigma[\boldsymbol{k}]^{2c[\boldsymbol{k}]}}. \tag{21}$$

The sets $\mathcal{B}_1$ and $\mathcal{B}_2$ are defined as

$$\mathcal{B}_1 \triangleq \left\{ \boldsymbol{k} \in \mathbb{N}^2 \,\middle|\, |X[\boldsymbol{k}]| \leq (K_e - 1)\alpha[\boldsymbol{k}] \right\}$$
$$\mathcal{B}_2 \triangleq \left\{ \boldsymbol{k} \in \mathbb{N}^2 \,\middle|\, (K_e - 1)\alpha[\boldsymbol{k}] < |X[\boldsymbol{k}]| \leq (K_e + 1)\alpha[\boldsymbol{k}] \right\}.$$

We have also defined the 2–D sequence

$$r'[\boldsymbol{k}] \triangleq \begin{cases} r[\boldsymbol{k}], & \text{if } Y[\boldsymbol{k}] \leq K_e\alpha[\boldsymbol{k}] \\ 0, & \text{otherwise} \end{cases} \tag{22}$$

whose mean and variance are

$$E[r'[\boldsymbol{k}]] = \tfrac{1}{2} \left(|\,|X[\boldsymbol{k}]| - 2\alpha[\boldsymbol{k}]|^{c[\boldsymbol{k}]} - |X[\boldsymbol{k}]|^{c[\boldsymbol{k}]} \right)$$
$$\operatorname{Var}(r'[\boldsymbol{k}]) = \tfrac{1}{4} \left(|\,|X[\boldsymbol{k}]| - 2\alpha[\boldsymbol{k}]|^{c[\boldsymbol{k}]} - |X[\boldsymbol{k}]|^{c[\boldsymbol{k}]} \right)^2.$$

The mean and variance of $r[\boldsymbol{k}]$ are given by (16) and (17). In Section VI-D, in which we discuss the experimental results, we will see how the modified statistic just derived can lead to a considerable improvement in the performance of the hidden information decoder.

C. Estimation of Distribution Parameters

In the preceding sections, we have assumed that the parameters $\sigma^2[\boldsymbol{k}]$ and $c[\boldsymbol{k}]$ of the distributions of the DCT coefficients were known. However, in practice these parameters must be estimated from the image intended to be tested $Z[\boldsymbol{k}]$. In other words, in a practical implementation, a parameter estimation stage has to be added at the input of the block devoted to the computation of sufficient statistics in Fig. 2.

As a first approach, the results presented in this paper correspond to a decoder structure in which $c[\boldsymbol{k}]$ is fixed to a unique value, independent of the image under test, for all the coefficients. Then, different values have been tried in order to analyze the impact of this parameter on the performance of the decoder
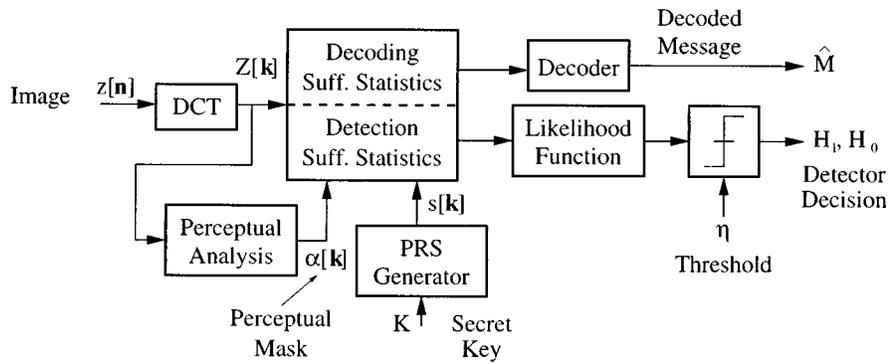
Fig. 2.   Block diagram of the watermark verification process.

for different images. For simplicity purposes, the variance $\sigma^2[\boldsymbol{k}]$ is estimated for each DCT coefficient using the estimator

$$\hat{\sigma}^2(i,\,j) = \hat{\sigma}_Y^2(i,\,j) - \frac{1}{N_b}\sum_{\boldsymbol{k}} \alpha_{i,\,j}^2[\boldsymbol{k}] \qquad (23)$$

where $\alpha_{i,\,j}[\boldsymbol{k}] \triangleq \alpha[8k_1 + i,\, 8k_2 + j]$ is the 2–D sequence that results after taking the values of the perceptual mask associated with the DCT coefficient $(i,\,j)$, N_b is the number of 8×8 pixel blocks that the image contains and

$$\hat{\sigma}_Y^2(i,\,j) = \frac{1}{N_b}\sum_{\boldsymbol{k}} X_{i,\,j}^2[\boldsymbol{k}] - \frac{1}{N_b^2}\left(\sum_{\boldsymbol{k}} X_{i,\,j}[\boldsymbol{k}]\right)^2. \quad (24)$$

The use of this variance estimator in the computation of the sufficient statistics introduces an additional amount of dependence between them and the watermark $W[\boldsymbol{k}]$, a dependence that is difficult to analyze theoretically. Hence, a slight deviation between the theoretical approximations and the empirical results will be observed due to this effect. Experimental data obtained from simulations show, anyway, that the error is negligible (see Section VI-D).

The problem of obtaining good values of c is out of the scope of this paper, since our main purpose is to illustrate how extending the correlating receiver structure to one based on a generalized Gaussian noise model leads to great benefits in terms of performance. However, some possible approaches to tackle with this problem in a practical implementation are now outlined.

One possible choice of c is to use a *constant* value, regardless of the image under analysis. In this case, an appropriate value of c should be chosen so that the "average" performance over a set of test images is clearly improved with respect to the correlating receiver structure. A good candidate as a fixed image independent parameter is $c = 0.8$, since this value has been reported to be the best choice for modeling non-DC DCT coefficients in the low frequency band [25].

It is possible, however, to adaptively estimate the value of c either for the ensemble of the ac DCT coefficients, or independently for each coefficient, given a certain image to be tested. This approach relies on the assumption that the DCT coefficients of a watermarked image can still be modeled as generalized Gaussian random variables with approximately the same value of c as in the original image. Estimates of c can be obtained by matching the sample mean absolute value and the

sample variance of the DCT coefficients to those of the generalized Gaussian distribution, as proposed in [23] and [26], by solving

$$c = F^{-1}\left(\frac{E[|X_{i,j}|]}{\sigma_{i,j}}\right) \qquad (25)$$

where

$$F(c) = \frac{\Gamma(2/c)}{\sqrt{\Gamma(1/c)\Gamma(3/c)}}. \qquad (26)$$

We propose an alternative technique, consisting in computing the ML estimator for the c parameter. This ML estimate can be obtained by maximizing the log-likelihood function, which can be proved to be, for the $(i,\,j)$ DCT coefficient,

$$L(i,\,j) = -\frac{N_b}{c}\ln\left(\frac{1}{N_b}\sum_{\boldsymbol{k}} |x[\boldsymbol{k}]|^c\right) + N_b\left(\ln c - \ln\Gamma(1/c)\right)$$
$$- \frac{N_b}{c}\left(\ln c + 1\right). \qquad (27)$$

In a practical implementation, the maximization of this function can be performed by evaluating it for a finite number of values of c defined on a discrete grid, as dense as possible depending on the available computational power of the application. Note that all but the first term in (27) can be evaluated by means of a look-up table. Also note that even though the proposed procedures for estimating the c parameter are not equivalent to finding the value of c which optimizes the decoding or detection performance, they can be expected to lead to similar results. Preliminary work performed by the authors shows that values of c obtained with this technique are quite close to the values for which best performance in terms of bit error rate are achieved (see Section VI-D).

D. Experimental Results

To contrast the analytical results derived in previous sections with empirical data, we have performed experiments with the test images shown in Fig. 3. These images have been chosen considering their different features. In all the tests performed, the watermark has been embedded in the DCT coefficients indicated in Fig. 4, located in the midfrequency range. In Figs. 5–7, we show theoretical and empirical curves representing the *bit error rate* (BER), i.e., the average number of bit errors, as a

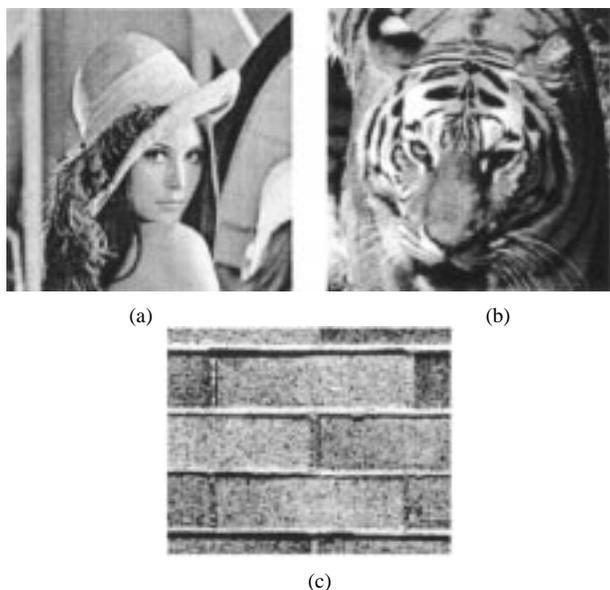(a)   (b)

(c)

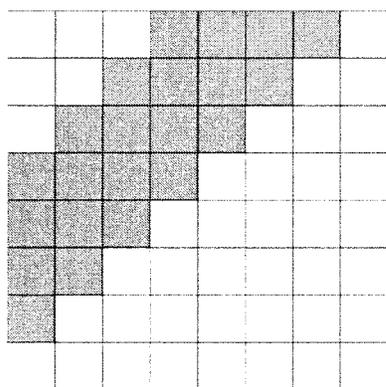Fig. 3. Test images used in the experiments. (a) Lena. (b) Tiger. (c) Brick.

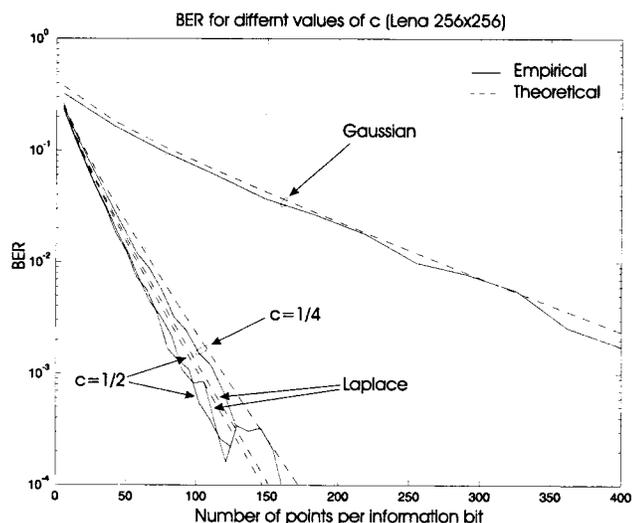Fig. 4. DCT coefficients altered by the watermark in the experiments.

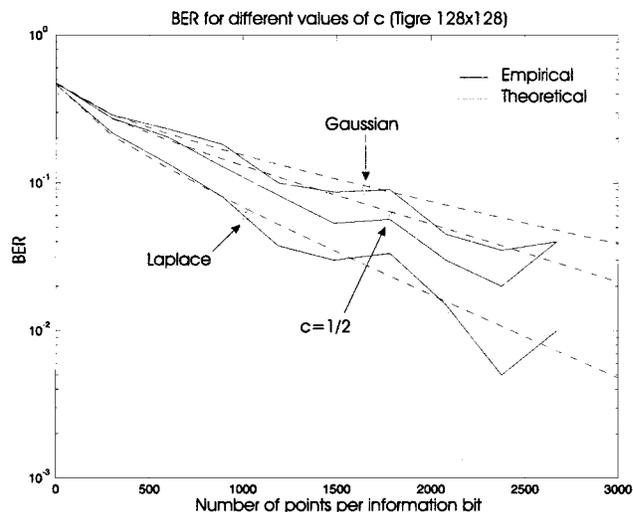Fig. 5. BER as a function of the pulse size for Lena (256×256).

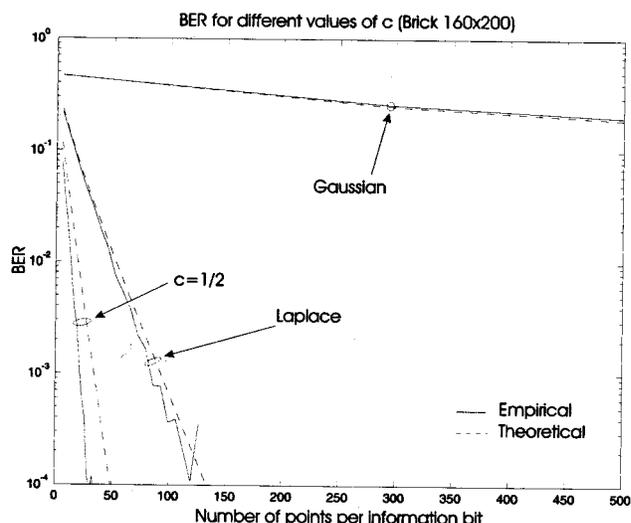Fig. 6. BER as a function of the pulse size for tiger (128×128).

Fig. 7. BER as a function of the pulse size for brick (160×200).

model are employed. The empirical measures have been obtained by averaging out over 100 keys randomly taken.

In all the tests we have used values of $c[k]$ that are identical for all the DCT coefficients, i.e., independent of k. We have tested several values of this parameter that are specially interesting, such as $c = 1$ (Laplace), $c = 2$ (Gaussian), and $c = 1/2$ (generalized Gaussian proposed in [22] as a reasonable model for low frequency DCT coefficients). The variances $\sigma^2[k]$, however, have always been estimated independently for each DCT coefficient from the watermarked image using the expressions in (23) and (24). With Lena and brick, we can see that the decoder achieving best performance is that associated to the generalized Gaussian model with $c = 1/2$. With tiger, however, the best decoder is the one derived from the Laplace model ($c = 1$). In any of the three test images we can see that a severe degradation in performance results when a Gaussian model ($c = 2$) is assumed. The detector associated with this value of c is actually equivalent to the commonly used detector structure, consisting in a correlator followed by a hard decisor. Thus, we can see how

function of the number of pixels in the sets $\mathcal{S}_i$ when the sufficient statistics derived from the generalized Gaussian noise
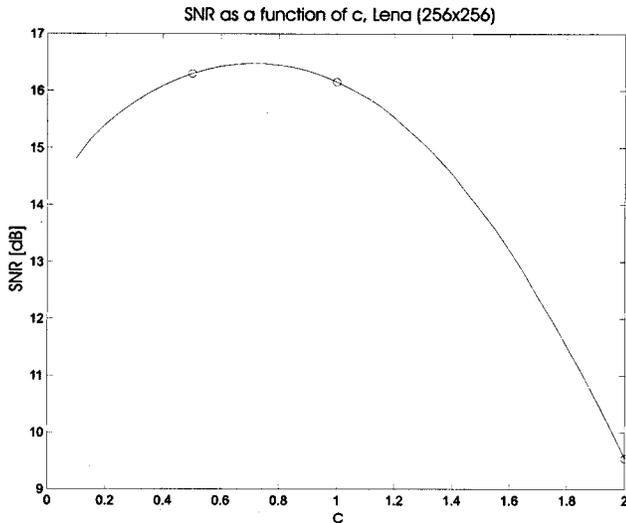
Fig. 8.   SNR as a function of c for Lena (256×256).
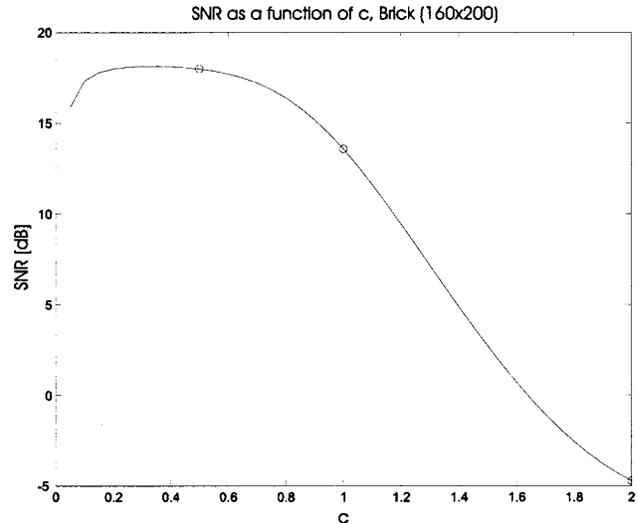
Fig. 10.   SNR as a function of c for brick (160×200).
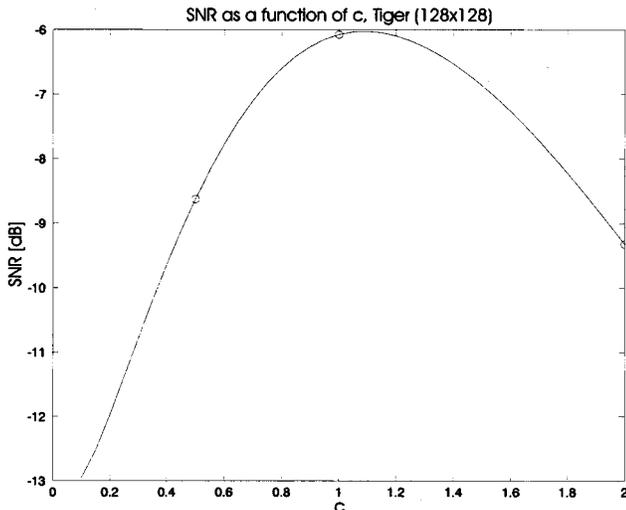
Fig. 9.   SNR as a function of c for tiger (128×128).

substantial improvements in the performance of the hidden information decoder can be achieved by just abandoning the pure Gaussian model, the only one employed so far in the literature on invisible watermarking and described as inadequate by researchers working on statistical modeling of DCT coefficients for images [22], [23], [27]. We can also see that experimental results closely fit the theoretical curves (dashed lines).

In Figs. 8–10, we plot the signal to noise ratio (SNR) [see (18)] of the sufficient statistics r_i as a function of the parameter c assumed in the decoder for the three test images. This SNR is defined as $\mathrm{SNR} = 20\log_{10}(a/\sigma_n)$. The curves have been obtained from the theoretical expressions for $E[r_i]$ and $\mathrm{Var}(r_i)$, that have been shown to lead to accurate estimates of the BER. We have encircled the points of the curves corresponding to the three values of c represented in Figs. 5–7. We can see that all the curves have a maximum at a value of c slightly different for each image. In Lena, for example, the maximum is somewhere between $c = 1$ (Laplace) and $c = 1/2$, while in tiger, the maximum is close to $c = 1$ (Laplace). The most surprising case is

brick, for which the maximum is even below $c = 1/2$. We have to take into account that this image is originally encoded with JPEG, so the distributions of the DCT coefficients are actually discrete and this fact may introduce deviations in the signal to noise ratio of a statistic specially designed for a continuous distribution. In all the three cases studied, the value corresponding to the Gaussian assumption ($c = 2$) lies far apart from the maximum, and its signal to noise ratio is much worse than for any of the other two cases of interest.

We have estimated the values of c following the ML criterion as proposed in Section VI-C. In our tests, the log-likelihood function in (27) has been maximized, for each of the three images, over a discrete grid of c values between 0.25 and 3, with a step size of 0.25. The results are estimates of c equal to 0.5, 0.75, and 0.25, for Lena, tiger, and brick, respectively. If we examine Figs. 8–10, we can see that these estimates are fairly close to the values optimum in terms of SNR. Note that ML estimation of c is not equivalent to searching the value of this parameter which optimizes the BER of the decoder. We can see, however, that ML estimates of c are close enough to achieve good performance.

We have also studied the influence of JPEG compression on the performance, for different values of the parameter c assumed in the watermark decoder. In Fig. 11, we show the BER for the decoders with $c = 1/2$ and $c = 1$ (Laplace) and the Lena image when JPEG compression is performed after the watermark embedding process with different quality factor values. We also show a curve for the optimum decoder, not derived in this paper (see [28]) specifically designed taking into account the probability distributions of the quantized DCT coefficients. A more detailed analysis of the effects of JPEG compression, including theoretical results, can be found in [28].

Finally, we have performed experiments using the sufficient statistics with point elimination that we analyzed in Section VI-B. To illustrate the impact of point elimination on the performance of the decoder, we show in Fig. 12 curves representing the signal to noise ratio as a function of the parameter K_e defining the threshold above which the DCT coefficients
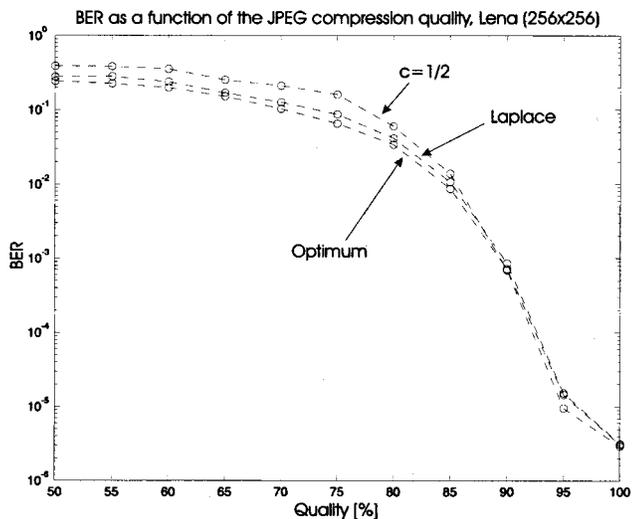
Fig. 11.   BER as a function of JPEG final quality for Lena (256×256).

Fig. 12.   SNR as a function of K_e for the point elimination scheme and the image Lena (256×256).
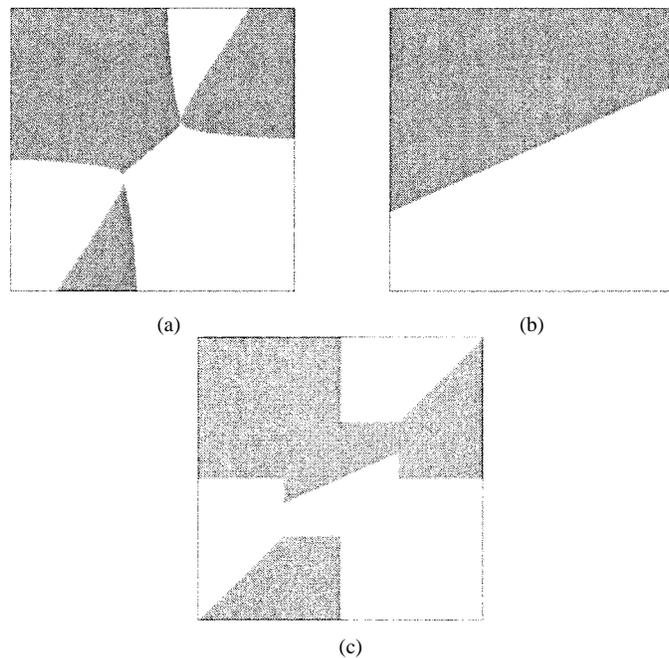

Fig. 13.   Decision regions for the watermark decoder in the 2–D case, with $\alpha_1 = \alpha_2 = 1, s_1 = s_2 = 1$ and $\sigma_1 = 1, \sigma_2 = 1.5$. (a) Generalized Gaussian, $c = 1/2$. (b) Gaussian. (c) Gaussian with elimination.
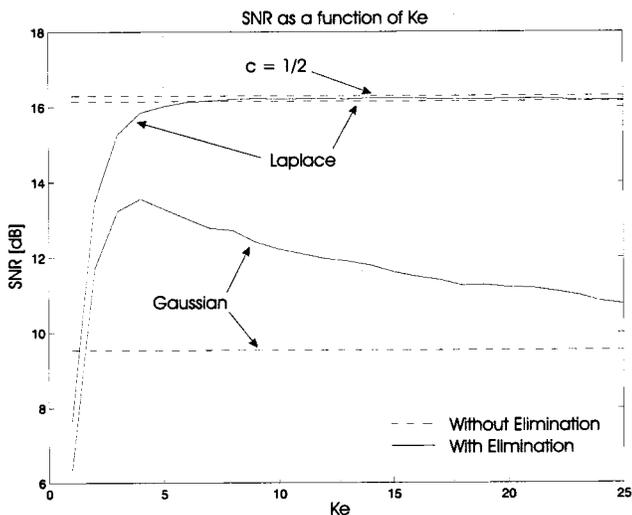
are ignored, for the Lena image. The dashed line represents the SNR for the three statistics (without point elimination) in which we have focused ($c = 1/2, 1, 2$). For the statistics with point elimination we have taken only the cases $c = 1$ and $c = 2$ in order to test for possible improvements (especially in the second case). We can see that the SNR with point elimination increases from $-\infty$ until it reaches a maximum, and then decreases until it converges to the SNR associated with the corresponding estimator without point elimination. Within a certain range of K_e values, the statistic with point elimination performs better than the corresponding statistic without elimination. We can clearly see that point elimination adds a considerable improvement when a Gaussian statistic is employed. The Laplace statistic without elimination is close to the best one, so point elimination hardly improves the SNR. Results from tests performed with the other two test images, not presented here, show a similar behavior.

An intuitive justification for the gains in performance obtained by using point elimination strategies can be given if we examine Fig. 13, in which we have represented graphically the decision regions in the 2–D case (i.e., only two DCT coefficients) assuming that only one information bit is encoded and that the pseudorandom sequence and the perceptual mask take the values $s_1 = s_2 = 1$ and $\alpha_1 = \alpha_2 = 1$, respectively. Assume that the correct value of c for the underlying distribution is $1/2$. In Fig. 13(a), we can see the optimum decision regions for this distribution. In Fig. 13(b), we have represented the decision regions associated with the correlating receiver structure ($c = 2$), while inFig. 13(c), we show the decision regions for a decoder with point elimination in which a value of $c = 2$ is assumed (correlating receiver). In the case (c), a threshold of $K_e = 2$ has been fixed for illustrative purposes. Also, to avoid indeterminate situations, when both DCT coefficients are above the threshold, only the one with largest amplitude is ignored. When the number of dimensions is high, the probability that all the coefficients are above the threshold is negligible, so this assumption will not affect the qualitative conclusions. As we can clearly see, the decision regions in (c) are much closer to the optimum than the decision regions of the correlating receiver without point elimination, so an improvement in the probability of error of the decoder can be expected.

Point elimination techniques are interesting for the design of practical implementations of the hidden information decoder. Instead of a detector that is well suited for certain images but requires complex computations, it is possible to use a detector based on a computationally simpler statistic, such as the ones derived from the Gaussian and Laplacian distributions, including a point elimination stage to improve the performance. However,

how to obtain the value of K_e that achieves the best gain in SNR for a given watermarked image remains an open line of research.

VII. Watermark Detection

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

$$H_1: Y[\mathbf{k}] = X[\mathbf{k}] + W[\mathbf{k}]$$
$$H_0: Y[\mathbf{k}] = X[\mathbf{k}] \tag{28}$$

where $X[\mathbf{k}]$ is the original image, not available during the test, and $W[\mathbf{k}]$ is a watermark generated from the secret key K that is tested. Hypothesis H_1 amounts to saying that the image under test contains a watermark generated with K, and H_0 amounts to saying that a watermark generated with K is not present in $Y[\mathbf{k}]$ (it might contain, though, a watermark generated with any other key). Note that it is not the goal of the watermark detection test to estimate the message in the watermark (if there is such a message) and that this task is left to the decoding stage. Therefore, during the design of the detector we must take into account the uncertainty about the value of the codeword vector $\mathbf{b}$. The optimum ML decision rule for the test formulated above is

$$\Lambda(Y) \underset{H_2}{\overset{H_1}{\gtrless}} \eta \tag{29}$$

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

$$\Lambda(Y) = \frac{1}{L} \sum_{l=1}^{L} \frac{f(Y|H_1, \mathbf{b_l})}{f(Y|H_0)}. \tag{30}$$

In the following section, we will derive the expression for this likelihood function when the original image $X[\mathbf{k}]$ is modeled by a generalized Gaussian distribution.

A. Optimum Detector for the Generalized Gaussian Model

If we assume that the coefficients of the original image $X[\mathbf{k}]$ follow the generalized Gaussian model studied in previous sections, then the log-likelihood function $l(Y) \overset{\Delta}{=} \ln \Lambda(Y)$ has the form

$$l(Y) = -\ln L + \sum_{\mathbf{k}} \beta[\mathbf{k}]^{c[\mathbf{k}]} |Y[\mathbf{k}]|^{c[\mathbf{k}]} + \ln \left(\sum_{l=1}^{L} \prod_{i=1}^{N} \right.$$
$$\left. \cdot \exp \left\{ -\sum_{\mathbf{k} \in \mathcal{S}_i} \beta[\mathbf{k}]^{c[\mathbf{k}]} |Y[\mathbf{k}] - b_{l,i}\alpha[\mathbf{k}]s[\mathbf{k}]|^{c[\mathbf{k}]} \right\} \right) \tag{31}$$

where $\beta[\mathbf{k}]$ is the parameter β in the generalized Gaussian pdf for the coefficient $X[\mathbf{k}]$, and can be obtained from $c[\mathbf{k}]$ and $\sigma[\mathbf{k}]$ using (6). In the special case (frequently considered) that the watermark does not carry hidden information, or in other words,

that there is only one pulse ($N = 1$) and it is modulated by a known value $b_1 = 1$, the likelihood function is

$$l(Y) = \sum_{\mathbf{k}} \beta[\mathbf{k}]^{c[\mathbf{k}]} \left(|Y[\mathbf{k}]|^{c[\mathbf{k}]} - |Y[\mathbf{k}] - \alpha[\mathbf{k}]s[\mathbf{k}]|^{c[\mathbf{k}]} \right). \tag{32}$$

We will measure the performance of the watermark detection test in terms of its *probability of false alarm P_F* and its *probability of detection P_D* for a given original image. The former is the proportion of keys for which we decide H_1 when the image is not watermarked, and the latter is the proportion of keys for which, when the image is watermarked, we correctly decide H_1. In Appendix B we show that the pdf's of $l(Y)$ conditioned to hypotheses H_1 and H_0 are approximately Gaussian with the same variance, σ_1^2, and means m_1 and $-m_1$, respectively, where

$$m_1 = \frac{1}{2} \sum_{\mathbf{k}} \beta[\mathbf{k}]^{c[\mathbf{k}]} \left(|X[\mathbf{k}] + \alpha[\mathbf{k}]|^{c[\mathbf{k}]} + |X[\mathbf{k}] - \alpha[\mathbf{k}]|^{c[\mathbf{k}]} \right)$$
$$- \sum_{\mathbf{k}} \beta[\mathbf{k}]^{c[\mathbf{k}]} |X[\mathbf{k}]|^{c[\mathbf{k}]} \tag{33}$$
$$\sigma_1^2 = \frac{1}{4} \sum_{\mathbf{k}} \beta[\mathbf{k}]^{2c[\mathbf{k}]}$$
$$\cdot \left(|X[\mathbf{k}] + \alpha[\mathbf{k}]|^{c[\mathbf{k}]} - |X[\mathbf{k}] - \alpha[\mathbf{k}]|^{c[\mathbf{k}]} \right)^2. \tag{34}$$

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

$$P_F = Q\left(\frac{\eta + m_1}{\sigma_1}\right), \quad P_D = Q\left(\frac{\eta - m_1}{\sigma_1}\right). \tag{35}$$

Let us define the SNR as $\text{SNR}_1 \overset{\Delta}{=} m_1^2/\sigma_1^2$. 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 (35), that

$$P_D = Q\left(Q^{-1}(P_F) - 2\sqrt{\text{SNR}_1}\right). \tag{36}$$

Hence, the *receiver operating characteristic* (ROC) 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.

B. Experimental Results

In order to measure the performance of the watermark detector presented in the previous section and to contrast the empirical data with the theoretical results, we have performed experiments with the test images Lena, tiger, and brick. In these tests we have used detector versions associated with the Laplacian, Gaussian, and generalized Gaussian with $c = 1/2$ distributions. In all cases the watermark embedding and detection processes have been applied to 1000 keys taken at random, only altering the DCT coefficients shown in Fig. 4.

In Tables I–III, we show the means and variances of the likelihood function $l(Y)$ conditioned to H_0 (m_0 and σ_0^2) and H_1

TABLE I
MEAN AND VARIANCE OF
THE LIKELIHOOD FUNCTION UNDER HYPOTHESIS H_0 AND H_1, FOR A
GENERALIZED GAUSSIAN DISTRIBUTION WITH $c = 1/2$

Image	m_0	m_1	σ_0^2	σ_1^2
Lena	-484.64	480.55	275.64	266.19
Tiger	-2.92	2.89	7.61	7.60
Brick	-409.43	409.24	8.67	8.45

TABLE II
MEAN AND VARIANCE OF THE LIKELIHOOD FUNCTION UNDER HYPOTHESIS
H_0 AND H_1, FOR A LAPLACE DISTRIBUTION

Image	m_0	m_1	σ_0^2	σ_1^2
Lena	-475.88	464.34	299.64	267.13
Tiger	-4.53	4.66	6.97	6.95
Brick	-116.28	116.22	11.31	11.02

TABLE III
MEAN AND VARIANCE OF THE LIKELIHOOD FUNCTION UNDER HYPOTHESIS
H_0 AND H_1, FOR A GAUSSIAN DISTRIBUTION

Image	m_0	m_1	σ_0^2	σ_1^2
Lena	-432.41	408.49	1492.87	1212.19
Tiger	-7.45	7.46	39.00	38.75
Brick	-17.16	16.78	71.29	70.78

(m_1 and σ_1^2), evaluated empirically from tests performed with 1000 keys for the three test images and c values of $1/2$, 1, and 2. We can see that the conclusions drawn in the theoretical analysis about the symmetry properties of the two distributions are correct, since m_1 and m_0 have approximately the same amplitude and opposite sign, and the variances σ_1^2 and σ_0^2 are approximately equal in all cases.

In Table IV, we can see both the theoretical and empirical "signal to noise ratios" SNR_1 for the three test images and the three detector structures studied. As we have already seen in the previous section, this parameter completely determines the shape of the ROC for each detector. First, notice that the empirical measurements fit the theoretical estimates. Besides this, the performance highly depends on the image characteristics in the DCT domain. See, for example, the large difference between the SNR_1 for tiger and the other two images. When we discussed the hidden information decoder, a similar situation appeared, since the probability of bit error for tiger was very high.

The shape of the ROC also depends on the value of c chosen for the detector. If we examine Table IV we can see that among the three values of c analyzed, the best performance for watermark detection is also achieved for the same values given in Section VI-D, since for Lena and brick the best value of c is $1/2$, while for tiger the best one is $c = 1$. Note, anyway, the remarkable fact that with *Tiger* the detector based on the Gaussian assumption performs better than the one with $c = 1/2$.

In order to verify graphically the precision of the theoretical estimates, we have plot in Fig. 14 the ROC associated with tiger for each of the three detectors.

VIII. CONCLUSIONS AND FURTHER WORK

In this paper, a family of DCT-domain watermarking techniques based on spread spectrum modulation schemes has been analyzed. Security and robustness are provided by means of a secret key whose value determines the output of the pseudorandom sequence generator and the pseudorandom sample permutation performed in the interleaving stage. The original image is not necessary during the watermark detection and decoding processes, for the sake of flexibility in the possible applications of the watermarking process. A statistical characterization of the original image's DCT coefficients has been provided, using the generalized Gaussian model. Then, optimal ML detector structures have been derived for both the watermark decoding (extraction) and watermark detection problems. Analytical approximations to performance measures conditioned to a given original image, such as the probability of error in decoding and the probabilities of false alarm and detection in watermark detection, have also been obtained. The analytical expressions allow us to measure beforehand the performance that can be expected for a given image and to analyze the influence that image characteristics and system parameters such as the hidden message length have on the final performance. They are also an invaluable help for the assignment of values to the decision threshold in the watermark detection test so that a maximum allowable probability of false alarm is guaranteed. Finally, all the analytical results have been contrasted with data obtained through experimentation. We have shown how by adequately modeling the decision problems involved in the ownership verification process in a statistical framework, and more precisely, by abandoning the Gaussian noise assumption usually made implicitly in the previous literature, substantial improvements in performance can be obtained.

In this paper, we have not studied a salient problem common to all watermarking techniques based on spread spectrum schemes and not requiring the availability of the original image during watermark extraction and detection: sensibility to geometrical transformations such as scaling and rotation. An iterative 2-D synchronization algorithm as that discussed in [29] is a possible solution to this problem, although its suitability for implementation in a real system still has to be studied by means of a theoretical analysis.

Channel coding schemes for error protection purposes can help in further improving the performance of the watermark decoder [30]. The analysis of the impact of channel codes on the performance and the derivation of the most appropriate coding schemes is left as a future line of research.

APPENDIX A
SUFFICIENT STATISTICS FOR WATERMARK DECODING

We will statistically characterize the coefficients r_i used in the watermark decoder assuming that the secret key is the only random variable in the watermarking system and that the original image $X[\boldsymbol{k}]$ is fixed. First, we will assume that no interleaving is present (see Section II) and calculate the mean and variance of r_i. For the sake of simplicity in the notation, let us define the 2–D sequence

$$r[\boldsymbol{k}] \triangleq |Y[\boldsymbol{k}] + \alpha[\boldsymbol{k}]s[\boldsymbol{k}]|^{c[\boldsymbol{k}]} - |Y[\boldsymbol{k}] - \alpha[\boldsymbol{k}]s[\boldsymbol{k}]|^{c[\boldsymbol{k}]}. \quad (37)$$

TABLE IV
EMPIRICAL AND THEORETICAL SIGNAL TO NOISE RATIO SNR$_1$ (IN dB)

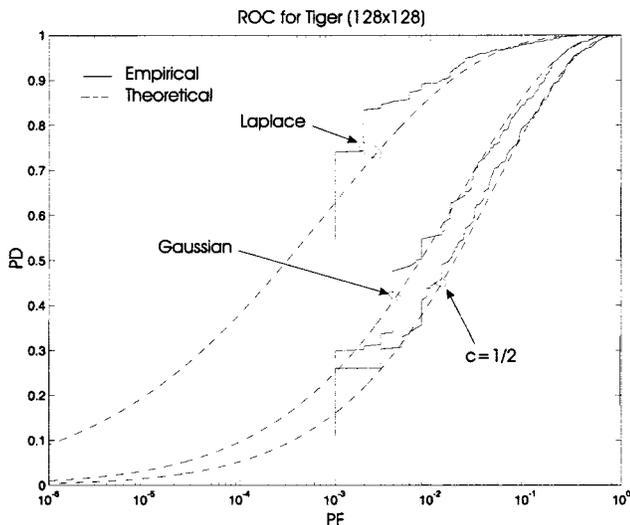| Image	$c = 1/2$		Laplace		Gaussian	
	Empirical	Theoretical	Empirical	Theoretical	Empirical	Theoretical
Lena	29.38	29.34	29.07	28.71	21.39	20.78
Tiger	0.41	0.41	4.95	4.65	1.57	1.65
Brick	42.97	43.23	30.89	30.81	6.00	6.26

Fig. 14.   Receiver operating characteristic (ROC) for tiger (128×128) for a pure watermark with no attacks.

The sufficient statistics r_i can be expressed as a function of this sequence as

$$r_i = \sum_{\boldsymbol{k} \in \mathcal{S}_i} \frac{r[\boldsymbol{k}]}{\sigma[\boldsymbol{k}]^{c[\boldsymbol{k}]}}. \tag{38}$$

Since each coefficient r_i is the sum of $|\mathcal{S}_i|$ independent random variables, we can infer, applying the central limit theorem, that these coefficients can be approximated by Gaussian random variables when the size of the modulation pulses is large enough. The mean and variance of this Gaussian distributions can be expressed as

$$E[r_i] = \sum_{\boldsymbol{k} \in \mathcal{S}_i} \frac{E[r[\boldsymbol{k}]]}{\sigma[\boldsymbol{k}]^{c[\boldsymbol{k}]}} \tag{39}$$

$$\mathrm{Var}(r_i) = \sum_{\boldsymbol{k} \in \mathcal{S}_i} \frac{\mathrm{Var}(r[\boldsymbol{k}])}{\sigma[\boldsymbol{k}]^{2c[\boldsymbol{k}]}}. \tag{40}$$

First, assume that $b_i = 1$. Then, $r[\boldsymbol{k}]$ can be expressed as

$$r[\boldsymbol{k}] = |X[\boldsymbol{k}] + 2\alpha[\boldsymbol{k}]s[\boldsymbol{k}]|^{c[\boldsymbol{k}]} - |X[\boldsymbol{k}]|^{c[\boldsymbol{k}]}. \tag{41}$$

In our experiments we have used sequences $s[\boldsymbol{k}]$ which can be modeled as an i.i.d. 2–D random process with a two-level discrete marginal distribution, $s[\boldsymbol{k}] \in \{-1, 1\}, \; \forall \boldsymbol{k}$. In this case, the mean and variance of $r[\boldsymbol{k}]$ are

$$E[r[\boldsymbol{k}]] = \tfrac{1}{2} \left[(|X[\boldsymbol{k}]| + 2\alpha[\boldsymbol{k}])^{c[\boldsymbol{k}]} + |\,|X[\boldsymbol{k}]| - 2\alpha[\boldsymbol{k}]|^{c[\boldsymbol{k}]} \right]$$
$$- |X[\boldsymbol{k}]|^{c[\boldsymbol{k}]} \tag{42}$$

$$\mathrm{Var}(r[\boldsymbol{k}]) = \tfrac{1}{4} \left[(|X[\boldsymbol{k}]| + 2\alpha[\boldsymbol{k}])^{c[\boldsymbol{k}]} - |\,|X[\boldsymbol{k}]| - 2\alpha[\boldsymbol{k}]|^{c[\boldsymbol{k}]} \right]^2. \tag{43}$$

If we substitute these expressions into (39) and (40), we can obtain the mean and variance of each sufficient statistic conditioned to $b_i = 1$, $i \in \{1, \cdots, N\}$. When $b_i = -1$, $r[\boldsymbol{k}]$ follows the same expression as in (41), with opposite sign. Therefore, the coefficient r_i will have the same mean, with opposite sign, and the same variance as those corresponding to $b_i = 1$. Furthermore, since the sets $\{\mathcal{S}_i\}_{i=1}^N$ are disjoint, and since $s[\boldsymbol{k}]$ is i.i.d., we can conclude that the coefficients r_i are uncorrelated.

If an interleaver is added for security purposes (see Section II), then the sets $\{\mathcal{S}_i\}_{i=1}^N$ will be key-dependent, so they must be modeled statistically and considered in the characterization of the coefficients r_i. We will assume for this purpose that the pseudorandom permutation of the signal $b[\boldsymbol{k}]$ (Section II) is performed in such a way that each DCT coefficient is assigned to any of the codeword coefficients b_i, $i \in \{1, \cdots, N\}$ with the same probability and that assignments for different DCT coefficients are approximately independent. So far we have analyzed the mean and variance of the sufficient statistics r_i conditioned to a certain tiling $\mathcal{T} = \{\mathcal{S}_i\}_{i=1}^N$. Using basic properties of the expectation operator, we know that

$$E[r_i] = E_{\mathcal{T}}[E[r_i|\mathcal{T}]] \tag{44}$$
$$\mathrm{Var}(r_i) = E_{\mathcal{T}}[\mathrm{Var}(r_i|\mathcal{T})] + \mathrm{Var}_{\mathcal{T}}(E[r_i|\mathcal{T}]). \tag{45}$$

The mean of r_i can be computed using the expression

$$E[r_i] = \sum_{\boldsymbol{k}} \frac{E[r[\boldsymbol{k}]]}{\sigma[\boldsymbol{k}]^{c[\boldsymbol{k}]}} \, \mathrm{Pr}\{\boldsymbol{k} \in \mathcal{S}_i\} = \frac{1}{N} \sum_{\boldsymbol{k}} \frac{E[r[\boldsymbol{k}]]}{\sigma[\boldsymbol{k}]^{c[\boldsymbol{k}]}}. \tag{46}$$

The first term in (45) can be expressed as

$$E_{\mathcal{T}}[\mathrm{Var}(r_i|\mathcal{T})] = \sum_{\boldsymbol{k}} \frac{\mathrm{Var}(r[\boldsymbol{k}])}{\sigma[\boldsymbol{k}]^{2c[\boldsymbol{k}]}} \, \mathrm{Pr}\{\boldsymbol{k} \in \mathcal{S}_i\}$$
$$= \frac{1}{N} \sum_{\boldsymbol{k}} \frac{\mathrm{Var}(r[\boldsymbol{k}])}{\sigma[\boldsymbol{k}]^{2c[\boldsymbol{k}]}}. \tag{47}$$

To analyze the second term in the right-hand side of (45), we first obtain

$$
\begin{aligned}
E_{\mathcal{T}}[E^2[r_i|\mathcal{T}]] &= \sum_{\boldsymbol{k}} \sum_{\boldsymbol{l}} \frac{E[r[\boldsymbol{k}]]E[r[\boldsymbol{l}]]}{(\sigma[\boldsymbol{k}]\sigma[\boldsymbol{l}])^{c[\boldsymbol{k}]}} \Pr\{\boldsymbol{k} \in \mathcal{S}_i, \boldsymbol{l} \in \mathcal{S}_i\} \\
&= \frac{N-1}{N^2} \sum_{\boldsymbol{k}} \frac{E^2[r[\boldsymbol{k}]]}{\sigma[\boldsymbol{k}]^{2c[\boldsymbol{k}]}} + \left(\frac{1}{N} \sum_{\boldsymbol{k}} \frac{E[r[\boldsymbol{k}]]}{\sigma[\boldsymbol{k}]^{c[\boldsymbol{k}]}}\right)^2 \\
&= \frac{N-1}{N^2} \sum_{\boldsymbol{k}} \frac{E^2[r[\boldsymbol{k}]]}{\sigma[\boldsymbol{k}]^{2c[\boldsymbol{k}]}} + E_{\mathcal{T}}^2[E[r_i|\mathcal{T}]]. \quad (48)
\end{aligned}
$$

From this, we infer that

$$
\begin{aligned}
\operatorname{Var}_{\mathcal{T}}(E[r_i|\mathcal{T}]) &= E_{\mathcal{T}}[E^2[r_i|\mathcal{T}]] - E_{\mathcal{T}}^2[E[r_i|\mathcal{T}]] \\
&= \frac{N-1}{N^2} \sum_{\boldsymbol{k}} \frac{E^2[r[\boldsymbol{k}]]}{\sigma[\boldsymbol{k}]^{2c[\boldsymbol{k}]}}. \quad (49)
\end{aligned}
$$

Hence, the overall variance is

$$
\operatorname{Var}(r_i) = \frac{1}{N} \sum_{\boldsymbol{k}} \frac{\operatorname{Var}(r[\boldsymbol{k}])}{\sigma[\boldsymbol{k}]^{2c[\boldsymbol{k}]}} + \frac{N-1}{N^2} \sum_{\boldsymbol{k}} \frac{E^2[r[\boldsymbol{k}]]}{\sigma[\boldsymbol{k}]^{2c[\boldsymbol{k}]}}. \quad (50)
$$

These expressions for the first and second order moments, in addition to the Gaussian approximation, fully statistically characterize the coefficients r_i.

APPENDIX B
LIKELIHOOD FUNCTION FOR WATERMARK DETECTION

In this appendix, we will characterize statistically the likelihood function $l(Y)$ for each of the hypothesis H_1 and H_0, assuming that $s[\boldsymbol{k}]$ is the only random element in the watermarking system and the original image is fixed. When H_0 is true, we have that $Y[\boldsymbol{k}] = X[\boldsymbol{k}], \; \forall \boldsymbol{k}$. Therefore

$$
l(Y) = \sum_{\boldsymbol{k}} \beta[\boldsymbol{k}]^{c[\boldsymbol{k}]} \left(|X[\boldsymbol{k}]|^{c[\boldsymbol{k}]} - |X[\boldsymbol{k}] - \alpha[\boldsymbol{k}]s[\boldsymbol{k}]|^{c[\boldsymbol{k}]}\right) \quad (51)
$$

which is a sum of statistically independent terms. Hence, after applying the central limit theorem we can approximate $l(Y)$ by a Gaussian random variable. 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 the mean of $l(Y)$ is

$$
\begin{aligned}
E[l(Y)|H_0] &= \sum_{\boldsymbol{k}} \beta[\boldsymbol{k}]^{c[\boldsymbol{k}]} |X[\boldsymbol{k}]|^{c[\boldsymbol{k}]} - \frac{1}{2} \sum_{\boldsymbol{k}} \beta[\boldsymbol{k}]^{c[\boldsymbol{k}]} \\
&\quad \cdot \left(|X[\boldsymbol{k}] + \alpha[\boldsymbol{k}]|^{c[\boldsymbol{k}]} + |X[\boldsymbol{k}] - \alpha[\boldsymbol{k}]|^{c[\boldsymbol{k}]}\right). \quad (52)
\end{aligned}
$$

In order to compute the variance of $l(Y)$, let us first obtain the term

$$
\begin{aligned}
&l(Y) - E[l(Y)|H_0] \\
&= \frac{1}{2} \sum_{\boldsymbol{k}} \beta[\boldsymbol{k}]^{c[\boldsymbol{k}]} \left(|X[\boldsymbol{k}] + \alpha[\boldsymbol{k}]|^{c[\boldsymbol{k}]} + |X[\boldsymbol{k}] - \alpha[\boldsymbol{k}]|^{c[\boldsymbol{k}]}\right) \\
&\quad - \sum_{\boldsymbol{k}} \beta[\boldsymbol{k}]^{c[\boldsymbol{k}]} |X[\boldsymbol{k}] - \alpha[\boldsymbol{k}]s[\boldsymbol{k}]|^{c[\boldsymbol{k}]} \quad (53)
\end{aligned}
$$

from which, considering that $s[\boldsymbol{k}]$ is i.i.d., it can be easily shown that

$$
\begin{aligned}
&\operatorname{Var}(l(Y)|H_0) \\
&= \frac{1}{4} \sum_{\boldsymbol{k}} \beta[\boldsymbol{k}]^{2c[\boldsymbol{k}]} \left(|X[\boldsymbol{k}] + \alpha[\boldsymbol{k}]|^{c[\boldsymbol{k}]} - |X[\boldsymbol{k}] - \alpha[\boldsymbol{k}]|^{c[\boldsymbol{k}]}\right)^2. \\
&\quad (54)
\end{aligned}
$$

On the other hand, when H_1 is true we have that $Y[\boldsymbol{k}] = X[\boldsymbol{k}] + \alpha[\boldsymbol{k}]s[\boldsymbol{k}]$, so the likelihood function is in this case

$$
l(Y) = \sum_{\boldsymbol{k}} \beta[\boldsymbol{k}]^{c[\boldsymbol{k}]} \left(|X[\boldsymbol{k}] + \alpha[\boldsymbol{k}]s[\boldsymbol{k}]|^{c[\boldsymbol{k}]} - |X[\boldsymbol{k}]|^{c[\boldsymbol{k}]}\right). \quad (55)
$$

If we fix an index $\boldsymbol{k}$ and compute the two possible values that the expression inside the summation in (55) takes for $s[\boldsymbol{k}] = 1$ and $s[\boldsymbol{k}] = -1$, we will see that they are equal to those we would obtain by applying the expression inside the summation in (51), with opposite sign. Therefore, we can infer that $l(Y)$ conditioned to H_1 is approximately Gaussian with mean and variance

$$
E[l(Y)|H_1] = -E[l(Y)|H_0] \quad (56)
$$
$$
\operatorname{Var}(l(Y)|H_1) = \operatorname{Var}(l(Y)|H_0). \quad (57)
$$

REFERENCES

[1] W. Bender, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for data hiding," *IBM Syst. J.*, vol. 35, pp. 313–336, 1996.

[2] D. Gruhl and W. Bender, "Information hiding to foil the casual counterfeiter," in *Proc. Information Hiding Workshop*. Portland, OR, Apr. 1998.

[3] R. G. van Schyndel, A. Z. Tirkel, and C. F. Osborne, "A digital watermark," in *Proc. IEEE Int. Conference on Image Processing*, Austin, TX, 1994, pp. 86–89.

[4] N. Nikolaidis and I. Pitas, "Copyright protection of images using robust digital signatures," in *Proc. IEEE ICASSP'96*, 1996, pp. 2168–2171.

[5] M. D. Swanson, B. Zhu, and A. H. Tewfik, "Robust data hiding for images," in *Proc. IEEE Digital Signal Processing Workshop*, Loen, Norway, Sept. 1996, pp. 37–40.

[6] J. R. Smith and B. O. Comiskey, "Modulation and information hiding in images," in *Proc. Int. Workshop on Information Hiding*. Cambridge, UK, May 1996, pp. 207–226.

[7] J. Zhao and E. Koch, "Embedding robust labels into images for copyright protection," in *Proc. Int. Congress on Intellectual Property Rights for Specialized Information, Knowledge and New Technologies*, R. Oldenbourg, Ed., Vienna, Austria, Aug. 21–25, 1995, pp. 242–251.

[8] R. B. Wolfang and E. J. Delp, "A watermark for digital images," in *Proc. 1996 Int. Conf. Image Processing*, vol. 3, Lausanne, Switzerland, Sept. 1996, pp. 219–222.

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

[10] J. J. K. Ó Ruanaidh, W. J. Dowling, and F. M. Boland, "Watermarking digital images for copyright protection," in *Proc. Inst. Elect. Eng. Conf. Vision, Image and Signal Processing*, vol. 143, Aug. 1996, pp. 250–256.

[11] A. G. Borş and I. Pitas, "Image watermarking using DCT domain constraints," in *Proc. IEEE ICIP'96*, Lausanne, Switzerland, Sept. 1996, pp. 231–234.

[12] B. Tao and B. Dickinson, "Adaptive watermarking in the DCT domain," in *ICCASP'97*. Munich, Germany, Apr. 1997, pp. 2985–2988.

[13] A. Piva, M. Barni, F. Bartolini, and V. Cappellini, "DCT-based watermark recovering without resorting to the uncorrupted original image," in *Proc. IEEE ICIP'97*, Santa Barbara, CA, USA, Oct. 1997, pp. 520–523.

[14] G. C. Langelaar, J. C. A. van der Lubbe, and R. L. Lagendijk, "Robust labeling methods for copy protection of images," in *Proc. SPIE Conf. Electronic Imaging*, San Jose, CA, Feb. 1997.

[15] M. D. Swanson, B. Zhu, and A. H. Tewfik, "Transparent robust image watermarking," in *Proc. IEEE Int. Conf. on Image Processing*, Lausanne, Switzerland, Sept. 1996, pp. 211–214.

[16] J. R. Hernández and F. Pérez-González, "Statistical analysis of watermarking schemes for copyright protection of images," *Proc. IEEE, Special Issue Identification and Protection of Multimedia Information*, vol. 87, pp. 1142–1166, July 1999.

[17] G. K. Wallace, "The JPEG still picture compression standard," *IEEE Trans. Consumer Electron.*, vol. 38, pp. 18–34, Feb. 1992.

[18] A. J. Ahumada and H. A. Peterson, "Luminance-model-based DCT quantization for color image compression," *Proc. SPIE on Human Vision, Visual Processing, and Digital Display III*, vol. 1666, pp. 365–374, 1992.

[19] J. A. Solomon, A. B. Watson, and A. J. Ahumada, "Visibility of DCT basis functions: Effects of contrast masking," in *Proc. Data Compression Conf.*, Snowbird, UT, 1994, pp. 361–370.

[20] A. B. Watson, "Visual optimization of DCT quantization matrices for individual images," in *Proc. AIAA Computing in Aerospace 9*, San Diego, CA, 1993, pp. 286–291.

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

[22] R. J. Clarke, *Transform Coding of Images*, New York: Academic, 1985.

[23] K. A. Birney and T. R. Fischer, "On the modeling of DCT and subband image data for compression," *IEEE Trans. Image Processing*, vol. 4, pp. 186–193, Feb. 1995.

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

[25] N. Tanabe and N. Farvardin, "Subband image coding using entropy-coded quantization over noisy channels," Univ. Maryland, College Park, Tech. Rep., Computer Science Technical Report Series, Aug. 1989.

[26] S. G. Mallat, "'A theory of multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, p. 674, July 1989.

[27] R. C. Reininger and J. D. Gibson, "Distributions of the two-dimensional DCT coefficients for images," *IEEE Trans. Commun.*, vol. COMM-31, pp. 835–839, June 1983.

[28] J. R. Hernández, "Watermarking techniques for copyright protection of digital images," Ph.D. dissertation (in Spanish), Univ. Vigo, Oct. 1998.

[29] 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.

[30] J. R. Hernández, J. M. Rodríguez, and F. Pérez-González, "Improving the performance of spatial watermarking of images using channel coding," *Signal Process.*, to be published.

Juan R. Hernández (A'96) received the Ingeniero de Telecomunicación degree from the University of Vigo, Spain, in 1993, the M.S. degree in electrical engineering from Stanford University, Stanford, CA, in 1996, and the Ph.D. degree in telecommunications engineering from the University of Vigo, Spain, in 1998.

From 1993 to 1995, he was a member of the Department of Communication Technologies, University of Vigo, where he worked on hardware for digital signal processing and access control systems for digital television. Since 1996, he has been a Research Assistant in the same department. His research interests include digital communications and copyright protection in multimedia.

Martín Amado Castro received the Ingeniero de Telecomunicacíon degree from the University of Vigo, Spain, in 1999.

Since 1998, he has worked for Intelsis as a Network Management Engineer. His research interests include watermarking and digital communications.

Fernando Pérez-González (M'90) received the Ingeniero de Telecomunicación degree from the University of Santiago, Spain, in 1990 and the Ph.D. degree from the University of Vigo, Spain, in 1993, both in telecommunications engineering.

He joined the faculty of the School of Telecommunications Engineering, University of Vigo, as an Assistant Professor in 1990 and is currently an Associate Professor in the same institution. He has visited the University of New Mexico, Albuquerque, for different periods spanning ten months. His research interests lie in the areas of digital communications, adaptive algorithms, robust control, and copyright protection. He has been project manager of different projects concerned with digital television, both for satellite and terrestrial broadcasting. He is co-editor of the book *Intelligent Methods in Signal Processing and Communications* (Boston, MA: Birkhauser, 1997) and has been Guest Editor of a special section of the EURASIP journal *Signal Processing* devoted to signal processing for communications.