Correspondence

Optimum Decoding and Detection of Multiplicative Watermarks

Mauro Barni, Franco Bartolini, Alessia De Rosa, and Alessandro Piva

***Abstract*—This work addresses the problem of optimum decoding and detection of a multibit, multiplicative watermark hosted by Weibull-distributed features: a situation which is classically encountered for image watermarking in the magnitude-of-DFT domain. As such, this work can be seen as an extension of the system described in a previous paper, where the same problem is addressed for the case of 1-bit watermarking. The theoretical analysis is validated through Monte Carlo simulations. Although the structure of the optimum decoder/detector is derived in the absence of attacks, some experimental results are also presented, giving a measure of the overall robustness of the watermark when attacks are present.**

***Index Terms*—Multibit watermarking, multiplicative watermarking, optimum decoding, watermark presence assessment.**

I. Introduction

In this work, we address the problem of optimum decoding and detection of a multiplicative, multibit watermark embedded in the magnitude of the DFT coefficients of the host image. Such a problem comes down to the decoding/detection of a watermark hosted by a set of features following a Weibull probability density function (pdf). The Weibull pdf, in fact, has recently been used to model the statistical behavior of the magnitude of DFT coefficients of digital images [1].

The problem of the optimum recovery of a watermark embedded in Weibull-distributed features has been investigated in [1] and [2]; however, such works address 1-bit watermarking, where the detector only has to decide whether a given watermark is present in the image at hand or not. In contrast, in multibit watermarking, the decoder must extract the hidden information without knowing it in advance. Optimum decoding of a multibit watermark has been considered in [3] and [4], where channel coding is also taken into account. These works, however, refer to the additive watermarking of generalized-Gaussian distributed features (possibly describing an image watermarking scheme operating in the DCT domain) and can not be applied when a different embedding rule is used.

As to multibit watermark detection, the problem of assessing the presence of a multibit watermark is usually faced with heuristically by first estimating the hidden message and then verifying the presence of such a particular message. This approach, however, is not optimum since it tends to produce a high false detection rate since the detector always looks for the most likely sequence. The only attempt to theoretically derive the optimum detection strategy for a readable watermark is given in [3], where additive watermarking is considered. No such an attempt has been performed for the multiplicative case.

The specific watermarking algorithm considered here relies on the embedding of a spread spectrum watermark, which is amplitude modulated by an antipodal information string. A pseudo-random sequence that is uniformly distributed in $[-1,\ 1]$ is first generated and split into N_b chunks; then, each chunk is amplitude-modulated by multiplying it by $+1$ or -1, thus allowing the introduction of N_b information bits. The modulated sequence is casted into the magnitude of DFT coefficients belonging to the midportion of the frequency spectrum by following the multiplicative strategy described in [1].

By assuming equally probable information bits, optimum decoding reduces to maximum likelihood (ML) estimation problem, thus allowing us to derive the structure of the optimum decoder in closed form. As to detection, the problem is formulated as a statistical hypothesis testing problem, thus allowing the derivation of the optimum detection strategy, consisting of the comparison between a likelihood ratio function against a threshold to be set according to the Neyman–Pearson criterion [5]. Unfortunately, whereas the likelihood ratio can be expressed in closed form, the optimum threshold has to be determined experimentally by observing the answer of the detector to a set of test watermarks. In spite of this, results are good ones, thus validating the proposed approach and the underlying theoretical analysis.

Our analysis does not consider the possible presence of noise and visual masking. For sake of completeness, however, the overall performance of the proposed detector/decoder in a more realistic scenario, where attacks and perceptual masking are taken into account, are evaluated through experimental results.

II. Information Encoding and Watermark Embedding

The multibit watermarking algorithm considered in this paper is an extension of the 1-bit watermarking scheme described in [1]. Let $\mathbf{x} = \{x_1 \cdots x_n\}$ be the set of host features (i.e., the magnitude of a set of mid-frequency DFT coefficients of the host image), and let $\mathbf{m} = \{m_1 \cdots m_n\}$ be a pseudo-random sequence uniformly taking values in $[-1,\ 1]$. The marked set of features $\mathbf{y} = \{y_1 \cdots y_n\}$ is obtained by modifying the host DFT coefficients according to the following rule:

$$y_i = x_i + \gamma m_i x_i \tag{1}$$

where γ is a parameter controlling the watermark strength. In [1], an optimum detection algorithm, which permits a decision as to whether a given set of features $\mathbf{y}$ contains a given watermark $\mathbf{m}^*$ or not, is described.

Here, we extend the system presented in [1] to obtain a multibit watermark. To be specific, the watermark payload is increased by splitting the watermark sequence $\mathbf{m}$ into N_b chunks, where N_b is the number of bits to be embedded. Then, each chunk is multiplied by $+1$ or -1 according to the information bit to be transmitted. The embedding rule expressed in (1) must now be modified as follows:

$$y_i = x_i + \gamma w_i x_i \tag{2}$$

where the amplitude-modulated watermark $\mathbf{w}$ is achieved by modulating the pseudo-random sequence $\mathbf{m}$ by means of N_b information

Manuscript received February 4, 2002; revised November 19, 2002. This work was supported by the Italian Space Agency (ASI) under Grant I/R/178/02 "Watermarking techniques for authenication and copyright protection of remote sensing images accessible through public and private thematic networks." The associate editor coordinating the review of this paper and approving it for publication was Prof. Pierre Moulin.

M. Barni is with the Department of Information Engineering, University of Siena, Siena, Italy (e-mail: barni@dii.unisi.it).

F. Bartolini and A. De Rosa are with the Department of Electronics and Telecommunications, University of Florence, Firenze, Italy (e-mail: barto@lci.det.unifi.it; derosa@lci.det.unifi.it).

A. Piva is with National Inter-University Consortium for Telecommunications (CNIT), University of Florence, Firenze, Italy (e-mail: alessandro.piva@cnit.it).

Digital Object Identifier 10.1109/TSP.2003.809371

bits $\mathbf{b} = \{b_1 \cdots b_{N_b}\}$, which assume value $+1$ for bit 1, and -1 for bit 0:

$$w_i = m_i b_k, \qquad i = 1, \ldots, n; \quad k = \left\lceil \frac{iN_b}{n} \right\rceil. \tag{3}$$

Note that the same bit b_k is used for all of the coefficients in the same watermark chunk. In the following, we will indicate the number of coefficients assigned to each bit by r.

By following the approach used in [1], the DFT region $\mathcal{S}$ hosting the watermark consists of the coefficients belonging to the low-medium frequencies of the spectrum so that invisibility and robustness are ensured at the same time.

In order to increase the robustness of the system against attacks affecting only a part of the frequency spectrum (e.g., bandpass filtering), the set of DFT samples assigned to each bit is chosen at random, i.e., the set of host coefficients is randomly partitioned into N_b nonoverlapping subsets $\{\mathcal{S}_k\}_{k=1}^{N_b}$.

III. Optimum Watermark Decoding

Given the embedding rule described above, we now look for the optimum watermark decoder. The goal is to propose a criterion that maximizes the probability of a correct decision or, equivalently, that minimizes the probability of error. To do so, let us denote by $\mathcal{C}_m$ the mth decision region, i.e., the set of points in the observed feature space that result in the decoding of the mth bit sequence $\mathbf{b}_m$. By assuming that all the possible 2^{N_b} sequences are equally probable, minimization of the error probability boils down to a maximum-likelihood (ML) optimum criterion, where the decoded sequence is obtained by looking for the sequence that maximizes $f_{\mathbf{y}}(\mathbf{y}|\mathbf{m},\ \mathbf{b})$, i.e.,

$$\hat{\mathbf{b}} = \arg \max_{l=1\cdots 2^{N_b}} f_{\mathbf{y}}(\mathbf{y}|\mathbf{m},\ \mathbf{b}_l) \tag{4}$$

where $\mathbf{y} = \{y_1 \cdots y_n\}$ indicates the set of observed host features, and $f_{\mathbf{y}}(\mathbf{y}|\mathbf{m},\ \mathbf{b}_m)$ is the pdf of the random vector $\mathbf{y}$ conditioned to the events $\mathbf{m}$ and $\mathbf{b}_m$. Assuming that both bits of the information sequence and the coefficients in $\mathbf{m}$ are independent of each other, and by assuming that the host DFT coefficients are independent as well, the previous equation can be written as

$$\hat{\mathbf{b}} = \arg \max_{l=1\cdots 2^{N_b}} \prod_{k=1}^{N_b} f_{\mathbf{y}_k}(\mathbf{y}_k|\mathbf{m}_k,\ b_{l_k}) \tag{5}$$

where $\mathbf{y}_k$ is the set of DFT coefficients hosting the kth bit, i.e., those coefficients belonging to $\mathcal{S}_k$, and $\mathbf{m}_k$ is the corresponding set of coefficients of the random sequence $\mathbf{m}$. By assuming that channel coding is not used, bit-wise decoding of the transmitted sequence can be performed without losing optimality. Under this assumption, the optimum decision criterion for the kth bit can be formulated as

$$\begin{aligned} \hat{b}_k &= \arg \max_{b_k \in \{-1,+1\}} f_{\mathbf{y}_k}(\mathbf{y}_k|\mathbf{m}_k,\ b_k) \\ &= \arg \max_{b_k \in \{-1,+1\}} \prod_{i \in \mathcal{S}_k} f_y(y_i|m_i,\ b_k) \end{aligned} \tag{6}$$

where we have exploited the knowledge that a given DFT coefficient y_i depends only on the corresponding watermark component.

Due to (1), the pdf $f_y(y)$ of a marked coefficient y_i subject to a watermark value $m_i b_k$ can be written as

$$f_y(y_i|m_i,\ b_k) = \frac{1}{1+\gamma m_i b_k} f_x\left(\frac{y_i}{1+\gamma m_i b_k}\right) \tag{7}$$

where $f_x(x)$ indicates the pdf of the original, nonmarked, magnitude of the DFT coefficient. According to previous studies [1], [2], we decided to model the magnitude of DFT coefficients through a Weibull pdf $f_W(x)$, which is defined as

$$f_W(x) = \frac{\beta}{\alpha}\left(\frac{x}{\alpha}\right)^{\beta-1} \exp\left[-\left(\frac{x}{\alpha}\right)^{\beta}\right] \quad x > 0 \tag{8}$$

where $\alpha > 0$ and $\beta > 0$ are real-valued positive constants controlling the pdf mean, variance, and shape. By inserting the above expressions into (6) and by adopting a logarithmic formulation, we obtain the following decision rule:

$$\hat{b}_k = \text{sign}\left[\sum_{i \in \mathcal{S}_k} \frac{(1+\gamma m_i)^{\beta_i} - (1-\gamma m_i)^{\beta_i}}{\alpha_i^{\beta_i}(1+\gamma m_i)^{\beta_i}(1-\gamma m_i)^{\beta_i}} y_i^{\beta_i} + \sum_{i \in \mathcal{S}_k} \beta_i \ln \frac{1-\gamma m_i}{1+\gamma m_i}\right] \tag{9}$$

where α_i and β_i indicate the shape parameters of the Weibull pdf modeling the ith feature sample. Equation (9) can be expressed in the following compact form:

$$\hat{b}_k = \begin{cases} +1, & \text{if } \displaystyle\sum_{i \in \mathcal{S}_k} v_i y_i^{\beta_i} > T_z \\ -1, & \text{otherwise} \end{cases} \tag{10}$$

where we let:

$$v_i = \frac{(1+\gamma_i m_i)^{\beta_i} - (1-\gamma m_i)^{\beta_i}}{\alpha_i^{\beta_i}(1+\gamma m_i)^{\beta_i}(1-\gamma m_i)^{\beta_i}} \tag{11}$$

and

$$T_z = \sum_{i \in \mathcal{S}_k} \beta_i \ln \frac{1+\gamma m_i}{1-\gamma m_i}. \tag{12}$$

Implementation of the optimum decoder requires that coefficients α_is and β_is are known. According to our implementation, such values are estimated directly on the watermarked image, trusting that for $\gamma \ll 1$, the watermark presence does not bias significantly the results of the estimate. In Table I, the values of α_is and β_is estimated on the *Lena* image are given. Such parameters have been obtained by subdividing the region of the frequency spectrum hosting the watermark into 16 subparts, inside which α_i and β_i have been assumed to be constant (see [1] for a more detailed description of how the frequency spectrum is split to estimate α_i and β_i).

Having derived the optimum decoder structure, we should now calculate the bit error probability in the absence of attacks. To this aim, a common approach consists of applying the central limit theorem assuming that $z = \sum v_i y_i^{\beta_i}$ approximately follows a normal distribution, whose mean and variance under the hypotheses that $b = 1$ and $b = -1$ can be easily calculated once β_is are known. For example, this approach is used in [1] and [3]. A problem with the analysis based on the central limit theorem is that the normal approximation rapidly becomes inadequate when the error probabilities to be estimated get increasingly smaller. This is exactly the case with digital watermarking, where error probabilities as low as 10^{-6} or 10^{-8} are easily encountered. To avoid

TABLE I
VALUES OF α_iS AND β_iS FOR THE *LENA* IMAGE. FOR AN EXACT DEFINITION OF THE SUBREGIONS USED TO ESTIMATE α_iS AND β_iS, SEE [1]

Subregion	α	β	Subregion	α	β
1	0.056	1.82	9	0.057	2.01
2	0.069	1.79	10	0.056	2.04
3	0.059	1.79	11	0.040	1.88
4	0.028	1.79	12	0.027	1.82
5	0.030	1.88	13	0.033	1.85
6	0.043	1.88	14	0.032	1.85
7	0.033	1.63	15	0.024	1.85
8	0.015	1.85	16	0.016	1.92

the problems deriving from the normal approximation, we derived the bit error probability via Monte Carlo simulations. To do so, we fixed the number of coefficients used for each bit as $r = 300$ and estimated a set of typical α_is and β_is (see Table I). Then, we generated a large number of random samples drawn from a set of Weibull distributions having the desired shape parameters. Eventually, we used such samples to hide a set of randomly generated messages and used these watermarked coefficients to evaluate the quantities appearing in (10)–(12). These permitted us to and check whether the hidden message was correctly recovered or not. We repeated this procedure for several values of γ, finally obtaining the plot shown in Fig. 1. In the figure, the PSNR corresponding to each value of γ is given, where PSNR is defined as

$$\text{PSNR} = 10\log_{10}\left(\frac{255^2}{\varepsilon^2}\right) \tag{13}$$

where ε^2 indicates the mean square error between the marked and the original image. For each point of the curve, we considered a number of trials such that at least 100 errors were found to keep the statistical significance of simulations high while keeping the computational burden reasonably low (this also explains why the plot does not account for PSNR values lower than 45 dB).

The results given in the previous section must be compared with the actual bit error rate obtained when watermarking real images. In order to perform such a comparison, we carried out some experimental tests. All the experiments were carried out on 512×512 black and white images, with a watermark embedded in the magnitude of DFT coefficients belonging to the diagonals from the 80th to the 160th (see [1] for more details), for a total of 18 960 marked coefficients. We partition the 18 960-bit-long pseudo-random sequence into 64 subparts and modulated each subpart with one payload bit, thus inserting 64 information bits ($r = 296$ apart for the last bit, for which we let $r = 312$). The modulated sequence was then embedded into the host data with different values of γ. All the diagrams shown in the figures have been achieved by averaging the results obtained on 512 different pseudo random sequences and three test images (namely *Lena*, *Tiffany*, and *Lake*), each hosting 64 bits, for a total of about 10^5 hidden bits.

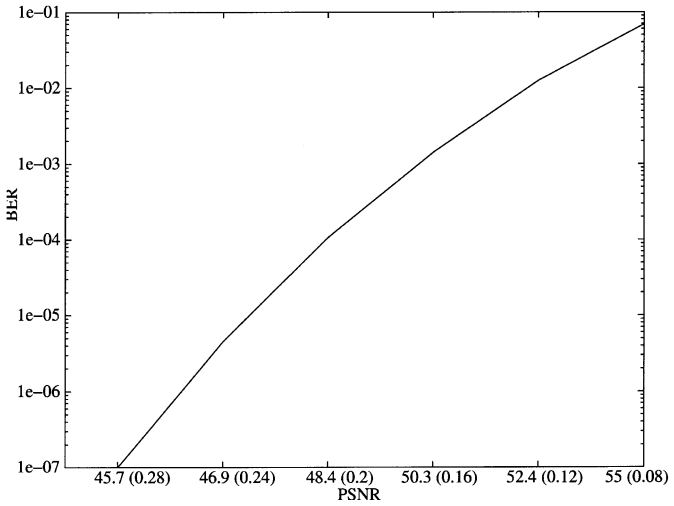

Fig. 1. Bit error probability obtained via Monte Carlo simulations. For each PSNR, the corresponding value of γ is given in round brackets. Results have been obtained by letting $r = 300$.

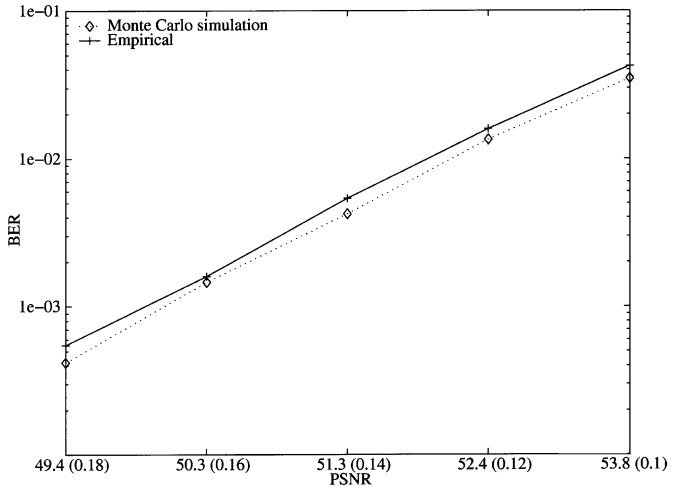

Fig. 2. Comparison between simulation results and empirical bit error probability, computed for different values of PSNR. The corresponding value of γ is given as well. Results have been obtained by letting $r = 300$.

In Fig. 2, the results achieved for different values of PSNR are plotted; as can be seen, the agreement with the model-based analysis is very good. Note that Fig. 2 covers a limited PSNR range to make the experimental analysis feasible.

IV. OVERALL PERFORMANCE

For sake of completeness, we also carried out some experiments to evaluate the performance of the proposed decoder in more realistic scenarios, i.e., when both attacks and visual masking are taken into account. Even if our analysis does not account for the presence of noise and visual masking, in fact, both these factors are likely to be present in many practical situations. For this set of experiments, we used a watermark strength $\gamma = 0.3$ (PSNR = 45 dB), that is the maximum allowable energy under the invisibility constraint when spatial masking is adopted. In the following, the results we obtained when JPEG compression, wavelet-based compression (JPEG2000), and median filtering are applied to the watermarked image are reported. As a first result, let us consider the robustness of the watermark against standard JPEG coding. The results we obtained are given in Fig. 3 (solid line), where the bit error rate is plotted as a function of the coding ratio expressed

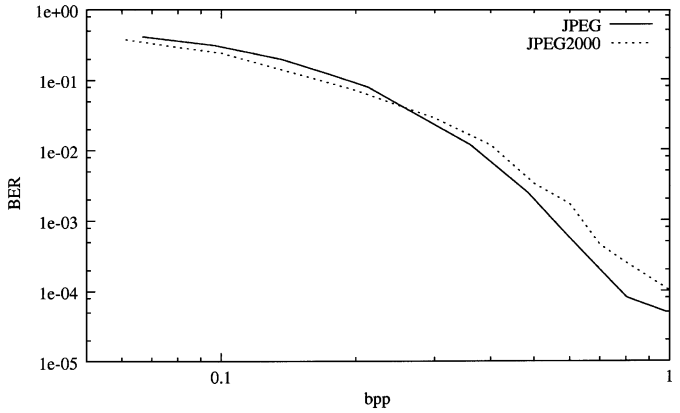

Fig. 3. Bit error probability in presence of JPEG (solid line) and JPEG 2000 (dashed line) coding with increasing bit per pixel.

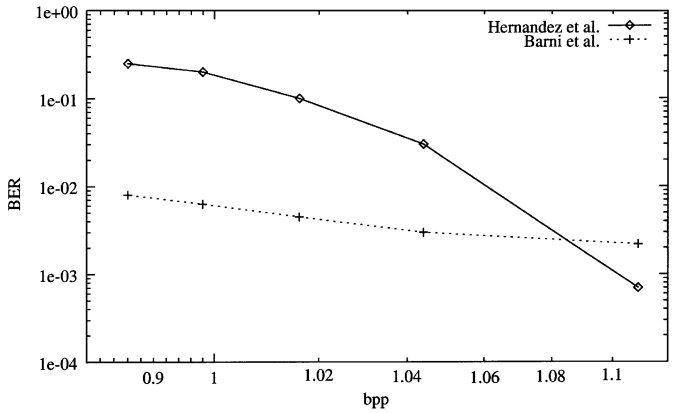

Fig. 4. Comparison between the performance of the proposed system and those reported in the work by Hernandez *et al.* Results refer to the watermarking of a 256×256 version of the *Lena* image ($r = 195$, $\gamma = 0.2$, PSNR $= 45$ dB) in the presence of JPEG coding with increasing bits per pixel. Results regarding Hernandez's systems have been directly taken from Hernandez's paper ($r = 220$, PSNR $= 45.1$ dB).

in bits per pixel. As it can be seen, results are rather good in that a bit error rate equal to 10^{-3} is now obtained for a compression ratio approximately equal to 15 (which is quite good if we consider that no error correction code is used).

Similar results are obtained in the JPEG2000 case (dashed line). Note again that performance can be significantly improved by applying channel coding protection [3], [4]. As to median filtering, results are comparable with those usually obtained in the related literature since the watermark is capable of surviving 3×3 median filtering ($P_e = 3 \cdot 10^{-4}$), but the BER increases significantly when larger filter windows are used.

We also compared the performance of our system with those obtained by the systems described in [3]. Such a system operates by embedding the watermark in the mid-frequency coefficients of block-DCT transform. In Fig. 4, the results obtained by Hernandez's system are contrasted with those of our system. Results refer to the watermarking of a 256×256 version of the *Lena* image hosting 24 information bits ($r = 195$ for our system, $r = 220$ for Hernandez's algorithm), in the presence of JPEG coding. The diagram referring to our system has been achieved by averaging the results obtained on 1000 different pseudo random sequences. In both cases, the watermarking strength was chosen so that PSNR $= 45$ dB. Upon inspection of the results, the validity of the proposed method comes out, especially at high compression ratios, where it outperforms the one by Hernandez *et al.*.

V. Watermark Presence Assessment

Blind verification of watermark presence for a multibit watermark is more difficult than in the 1-bit case. This is because the detector does not know the exact bit sequence hidden in the image.

In this paragraph, we derive the optimum detection structure for the watermarking system described in the previous sections. Optimality is based on the Neyman–Pearson criterion, i.e., we minimize the missed detection probability for a fixed false detection rate. The problem is formulated as a statistical hypothesis testing problem: The hypothesis H_1, that data contain a spreading sequence $\mathbf{m}^* = \{m_1^* \cdots m_n^*\}$, modulated by one of the 2^{N_b} possible bit sequences $\mathbf{b}$, is tested against the hypothesis H_0, that the data do not contain $\mathbf{m}^*$. The optimum decision criterion is based on the likelihood ratio $\ell(\mathbf{y})$ [1]:

$$\ell(\mathbf{y}) = \frac{f_{\mathbf{y}}(\mathbf{y}|\mathbf{m}^*)}{f_{\mathbf{y}}(\mathbf{y}|\mathbf{0})} \tag{14}$$

where $f_{\mathbf{y}}(\mathbf{y}|\mathbf{m}^*)$ denotes the pdf of the observed coefficients $\mathbf{y}$ conditioned to the presence of a given spreading sequence $\mathbf{m}^*$, and $f_{\mathbf{y}}(\mathbf{y}|\mathbf{0})$, which is the same pdf conditioned to the presence of the null sequence $\mathbf{0}$. In fact, as it has been proved in [1], if γ is reasonably small, the pdf of the observed coefficients conditioned on the absence of $\mathbf{m}^*$, $f_{\mathbf{y}}(\mathbf{y}|\mathbf{m} \neq \mathbf{m}^*)$, can be approximated by $f_{\mathbf{y}}(\mathbf{y}|\mathbf{0})$.

The pdf of the watermarked coefficients $f_{\mathbf{y}}(\mathbf{y}|\mathbf{m}^*)$ is obtained by integrating out the 2^{N_b} possible bit sequences. Using (2) and (3), $f_{\mathbf{y}}(\mathbf{y}|\mathbf{m}^*)$ can be rewritten as

$$\begin{aligned} f_{\mathbf{y}}(\mathbf{y}|\mathbf{m}^*) &= \prod_{k=1}^{N_b} f_{\mathbf{y}}(\mathbf{y}_k|\mathbf{m}_k^*) \\ &= \prod_{k=1}^{N_b} \{f_{\mathbf{y}}(\mathbf{y}_k|\mathbf{m}_k^*, -1)p(b_k = -1) \\ &\qquad + f_{\mathbf{y}}(\mathbf{y}_k|\mathbf{m}_k^*, +1)p(b_k = +1)\}. \end{aligned} \tag{15}$$

By assuming equal *a priori* probabilities, we get

$$f_{\mathbf{y}}(\mathbf{y}|\mathbf{m}^*) = \prod_{k=1}^{N_b} \tfrac{1}{2} \left\{ \prod_{i \in \mathcal{S}_k} f_y(y_i|m_i^*, -1) + \prod_{i \in \mathcal{S}_k} f_y(y_i|m_i^*, +1) \right\}. \tag{16}$$

By inserting the previous equation in (14) and using (7), the log likelihood ratio $\mathcal{L}(\mathbf{y})$ can be computed:

$$\begin{aligned} \mathcal{L}(\mathbf{y}) &= \ln \ell(\mathbf{y}) \\ &= \sum_{k=1}^{N_b} \left\{ -\ln 2 + \ln \left[\prod_{i \in \mathcal{S}_k} \left(\frac{1}{1 - \gamma m_i^*} \right)^{\beta_i} \right. \right. \\ &\qquad \cdot \exp \left[\left(\frac{y_i}{\alpha_i} \right)^{\beta_i} \left[1 - \left(\frac{1}{1 - \gamma m_i^*} \right)^{\beta_i} \right] \right] \\ &\qquad + \prod_{i \in \mathcal{S}_k} \left(\frac{1}{1 + \gamma m_i^*} \right)^{\beta_i} \\ &\qquad \left. \left. \cdot \exp \left[\left(\frac{y_i}{\alpha_i} \right)^{\beta_i} \left[1 - \left(\frac{1}{1 + \gamma m_i^*} \right)^{\beta_i} \right] \right] \right] \right\}. \end{aligned} \tag{17}$$

In order to specify completely the optimum detection criterion, the log likelihood ratio must now be compared against a threshold to decide if the given code $\mathbf{m}^*$ is present in the host data or not. By relying on the Neyman–Pearson criterion [5], the threshold is chosen in such a way that the missed detection rate is minimized subject to a fixed false

TABLE II
COMPARISON BETWEEN THE ACTUAL FALSE DETECTION RATE COMPUTED THROUGH MONTE CARLO SIMULATIONS (MC) AND THE ERROR RATE SET UNDER THE NORMALITY ASSUMPTION (CLT)

Target P_f (CLT)	Actual P_f (MC)
10^{-1}	$8.7 \cdot 10^{-2}$
10^{-2}	$7.1 \cdot 10^{-3}$
10^{-3}	$1.2 \cdot 10^{-3}$
10^{-4}	$4.3 \cdot 10^{-4}$
10^{-5}	$6.8 \cdot 10^{-5}$

positive probability, say $\overline{P_{FA}}$. Once $\overline{P_{FA}}$ has been fixed, the threshold λ can be computed by means of the relation:

$$\overline{P_{FA}} = P(\mathcal{L}(\mathbf{y}) > \lambda | H_0) = \int_{\lambda}^{+\infty} f_{\mathcal{L}}(\mathcal{L}|H_0)\, d\mathcal{L} \qquad (18)$$

where $f_{\mathcal{L}}(\mathcal{L}|H_0)$ is the pdf of $\mathcal{L}$ conditioned on H_0. We are now faced with the same difficulty we encountered in the decoding case: namely, derive a good estimate of $f_{\mathcal{L}}(\mathcal{L}|H_0)$. A first possibility consists of using Monte Carlo simulations to estimate the false detection probability for different values of λ and then choosing the threshold ensuring the desired false detection rate. Such an approach, however, is very computationally intensive, thus calling for a simpler, possibly suboptimal, solution. With these considerations in mind, we assume that $\mathcal{L}$ is normally distributed (the normality assumption can be supported by CLT arguments) and estimate its mean and variance by evaluating $\mathcal{L}$ for t fake spreading sequences $\{\mathbf{m}_i\}$, $1 \leq i \leq t$, i.e.,

$$\hat{\mu}_{\mathcal{L}} = \frac{1}{t} \sum_{i=1}^{t} \mathcal{L}_i \qquad (19)$$

$$\hat{\sigma}^2_{\mathcal{L}} = \frac{1}{t-1} \sum_{i=1}^{t} (\mathcal{L}_i - \hat{\mu}_{\mathcal{L}})^2 \qquad (20)$$

where, by $\mathcal{L}_i$, the log likelihood ratio corresponding to the ith fake spreading sequence is meant. Using this approximation for $f_{\mathcal{L}}$, we can evaluate $\overline{P_{FA}}$ in (18). Of course, the higher the t, the better the estimates of $\mu_{\mathcal{L}}$ and $\sigma^2_{\mathcal{L}}$. We found experimentally that a good tradeoff between computational complexity and accuracy of results can be obtained by letting $t = 100$. In order to evaluate the error introduced by the normality assumption, we compared the actual bit error rate computed through Monte Carlo simulations and the target bit error rate set by using the normal assumption. The results we obtained are given in Table II, where the actual bit error rate is compared with the target one. As can be seen, the error introduced by the normality assumption gets significant for low error rates. When using this simplified approach, then, it is necessary for the detector to compensate for such an error, e.g., by adopting a conservative threshold.

We also evaluated the performance of the simplified detector in the presence of JPEG/JPEG2000 coding and median filtering. While the presence of attacks and masking clearly introduces a discrepancy between theory and experiments, these results are very important since they give a good indication of detector effectiveness in real scenarios. The same experimental set up described previously was used to mark ten test images. Then, JPEG coding with increasing compression ratio was applied, and then, we searched for the watermark. We also looked for 100 fake watermarks to check whether a false alarm was generated or not. In the first column of Table III, the highest compression ratio (in bits per pixel) the watermark was able to survive (without generating a single false alarm) is given. The second column of Table III gives results for JPEG2000 coding. As for median filtering, we obtained the results illustrated in the third column of Table III. On average, reliable detection can be ensured for a window size up to 5×5: a result that is comparable with the best 1-bit watermarking schemes described in the literature.

TABLE III
WATERMARK DETECTION IN PRESENCE OF ATTACKS. FOR EACH ATTACK, THE HIGHEST ATTACK STRENGTH SURVIVED BY THE WATERMARK IS GIVEN (HIGHEST COMPRESSION RATIO IN BITS PER PIXEL, MAXIMUM WINDOW SIZE)

Image	JPEG	JPEG 2000	Median filtering
Baboon	0.2	0.14	5
Boat	0.2	0.18	5
Bridge	0.2	0.21	5
Couple	0.2	0.16	5
f16	0.2	0.19	3
Harbor	0.2	0.21	5
House	0.2	0.17	5
Lake	0.2	0.18	5
New York	0.3	0.21	5
Tiffany	0.2	0.15	5
Average	0.21	0.18	4.8

VI. CONCLUSIONS

We derived the structure of the optimum decoder for multiplicative, multibit watermarking of digital images in the magnitude-of-DFT domain. From this point of view, this work can be seen as an extension to multiplicative watermarking of the analysis carried out in [3] for the additive case. The analysis is given in [1] is extended as well since in such a work, only 1-bit watermarking was considered. Both simulations and experimental results witness the validity of the proposed approach, which is capable of ensuring a rather low BER, especially if we consider that significant improvements can be obtained by applying channel coding techniques to protect the embedded bit stream [4]. We also explored the possibility of distinguishing between a watermarked image and a nonwatermarked one in the multibit case. Even if the detection threshold could not be determined analytically, we proposed a suboptimal solution where the threshold is estimated directly on the image at hand. We validated this approach experimentally.

REFERENCES

[1] M. Barni, F. Bartolini, A. De Rosa, and A. Piva, "A new decoder for the optimum recovery of nonadditive watermarks," *IEEE Trans. Image Processing*, vol. 10, pp. 755–766, May 2001.

[2] J. Oostveen, T. Kalker, and J.-P. Linnartz, "Optimal detection of multiplicative watermarks," in *Proc. Tenth Eur. Signal Process. Conf.*, Tampere, Finland, Sept. 2000.

[3] J. R. Hernandez, M. Amado, and F. Perez-Gonzales, "DCT-domain watermarking techniques for still images: Detector performance analysis and a new structure," *IEEE Trans. Image Processing*, vol. 9, pp. 55–68, Jan. 2000.

[4] F. Perez-Gonzalez, J. R. Hernandez, and F. Balado, "Approaching the capacity limit in image watermarking: A perspective on coding techniques for data hiding applications," *Signal Process.*, vol. 81, no. 6, pp. 1215–1238, June 2001.

[5] J. V. Di Franco and W. L. Rubin, *Radar Detection*. Dedham, MA: Artech House, 1980.