Fixed-Distortion Orthogonal Dirty Paper Coding for Perceptual Still Image Watermarking

Andrea Abrardo and Mauro Barni

Department of Information Engineering, University of Siena
Via Roma 56, 53100 Siena, ITALY
{abrardo barni}@dii.unisi.it

Abstract. A new informed image watermarking technique is proposed incorporating perceptual factors into dirty paper coding. Due to the equi-energetic nature of the adopted codewords and to the use of a correlation-based decoder, invariance to constant value-metric scaling (gain attack) is automatically achieved. By exploiting the simple structure of orthogonal and Gold codes, an optimal informed embedding technique is developed, permitting to maximize the watermark robustness while keeping the embedding distortion constant. The maximum admissible distortion level is computed on a block by block basis, by using Watson's model of the Human Visual System (HVS). The performance of the watermarking algorithm are improved by concatenating dirty paper coding with a turbo coding (decoding) step. The validity of the assumptions underlying the theoretical analysis is evaluated by means of numerical simulations. Experimental results confirm the effectiveness of the proposed approach.

1 Introduction

Several digital watermarking methods trying to put into practice the hints stemming from the information-theoretic analysis of the watermarking game have been proposed. The main merit of these schemes, globally termed as informed watermarking algorithms, is that they permit to completely reject the interference between the cover signal and the watermark, thus leading to systems in which, in the absence of attacks, a zero error probability is obtained.

Random binning coding (or dirty paper coding) lies at the hearth of the informed watermarking approach [1]. To be specific, let us introduce an auxiliary source of randomness U, let B indicate the set with all the possible to-be-hidden messages, and let 2^{nR} be the number of messages contained in it. Finally, let C be the source emitting the cover feature sequence. The embedder first generates a codebook $\mathcal{U}$ consisting of 2^{nR_t} entries (call them $\mathbf{u}$'s) which are randomly generated so to span uniformly the set of typical sequences of $\mathcal{U}$ (for a tutorial introduction to typical sequences see [2, 3]). Then $\mathcal{U}$ is randomly (and uniformly) split into 2^{nR} bins (sub-codebooks) each containing $2^{n(R_t-R)}$ codewords. It is then possible to associate each message $\mathbf{b} \in B$ to a bin of $\mathcal{U}$. In order to transmit a message $\mathbf{b}$, the embedder looks at the host feature sequence $\mathbf{c}$ that is going to host the message, then an entry in the bin indexed by $\mathbf{b}$ is looked for which is jointly typical with $\mathbf{c}$. Next it maps the cover features $\mathbf{c}$ into a marked feature sequence $\mathbf{c}_w$ which is jointly typical with $\mathbf{u}$ and $\mathbf{c}$. At the other side, the decoder

J. Fridrich (Ed.): IH 2004, LNCS 3200, pp. 52–66, 2004.

receives a sequence $\mathbf{r}$. In order to estimate the transmitted message, the decoder looks for a unique sequence $\mathbf{u}^*$ in $\mathcal{U}$ which is jointly typical with $\mathbf{r}$ and outputs the message corresponding to the bin $\mathbf{u}^*$ belongs to. The decoder declares an error if more than one, or no such typical sequence exists. If R is lower than the watermarking capacity then it is possible to choose R_t so that the error probability averaged over al possible codes $\mathcal{U}$ tends to 0 as the length n of the transmitted sequence tends to infinity. The major problem with the random binning approach is that when n increases the dimension of the codebook becomes unmanageable, thus calling for the construction of structured codebooks allowing for an efficient search.

The most popular solution to put the random binning approach into practice is through the use of lattice based codebooks [4, 5, 6, 7]. The major weakness of the lattice approach, is that these schemes are vulnerable against constant value-metric scaling of the host features, a very common operation which consists in multiplying the host feature sequence by a constant factor g which is unknown to the decoder.

To overcome this problem, Miller et al. [8, 9] proposed to use equi-energetic codebooks and a correlation-based decoder, so that invariance to the presence of the constant gain g is automatically achieved. Their system relies on a dirty paper Trellis in which several paths are associated to the same message.

Of course equi-energetic codes do a much worse job in uniformly covering the host feature space, hence it is necessary to devise a particular embedding strategy which permits to move the host features sequence into a point within the decoding region associated to the to-be-transmitted message. This can be done either by fixing the watermark robustness and trying to minimize the embedding distortion, or by fixing the embedding distortion while maximizing the watermark robustness. In [8, 9], a sub-optimum, fixed-robustness, embedding strategy is proposed. In [10], the simple structure of orthogonal, and pseudo-orthogonal, codes is exploited to derive an optimum fixed-robustness embedding algorithm leading to performance which are superior to those obtained by Miller et al. with the further advantage of a reduced computational burden.

A difficulty with the fixed-robustness approach, is that the robustness constraint does not allow to take perceptual factors into account. As a matter of fact, in order to diminish the visibility[1] of the watermark, it is desirable that some features are marked less heavily than others, leading to a constraint on the maximum allowable distortion. In this paper, we extend the analysis contained in [10], to develop a fixed-distortion embedding algorithm for still image watermarking. Then we will use such an algorithm to incorporate perceptually driven considerations within the embedding step. Watermark embedding is performed in the block-DCT domain, since the Human Visual System (HVS) behavior is better modelled by working in the frequency domain. More specifically, we rely on the popular Watson's model [11, 12] measuring the maximum allowable distortion a block-DCT coefficient can sustain before the modification becomes visible. Watson's measure is used to constrain the maximum allowable embedding distortion on a block-by block basis.

Experiments and simulations were carried out to validate both the effectiveness of the proposed embedding strategy and to estimate the overall performance of the new watermarking system in terms of invisibility and robustness. In particular, the experi-

[1] We focus on still image watermarking.

ments demonstrated an excellent robustness against attacks involving scaling of the host features and a moderate robustness against more classical attacks such as noise addition and JPEG compression. Watermark invisibility was satisfactorily reached as well.

This paper is organized as follows. In section 2 the basic ideas behind dirty paper coding by means of orthogonal codes are reviewed. In section 3 the optimal algorithm for fixed-distortion embedding is derived, and the extension to quasi-orthogonal dirty paper coding presented. Section 4 explains how perceptual factors are incorporated within the fixed-distortion embedding scheme. The adoption of multistage (turbo) decoding to improve the overall performance of the system is described in section 5. Simulation and experimental results are presented in section 6. Finally, in section 7 some conclusions are drawn and directions for future research highlighted.

2 Orthogonal Dirty Paper Coding

In this section we briefly review the basic ideas of orthogonal dirty paper coding. For a more detailed analysis readers are referred to [10].

Let $\mathbf{c}$ represent the cover feature vector of length $n = 2^w$ and $\mathbf{U}$ a real $n \times n$ unitary matrix such as $\mathbf{U}^T\mathbf{U} = \mathbf{I}_n$ [2]. Each column of $\mathbf{U}$, say it $\mathbf{u}_i$, $i = 0, \ldots, n-1$, represents one out of n available codewords that can be associated to the information blocks to be embedded within $\mathbf{c}$. It is then assumed that a block of k bits is transmitted every side information block of length n and that each k-bit block is associated with one codeword which will be referred to as the carrier codeword. Note that, since the number of available codewords is n, a clear limit exists for k, i.e., $k \leq \log_2(n)$, or, equivalently, $k \leq w$.

Let now consider a partition of $\mathcal{U}$ into 2^k disjoint subsets Q_l, $l = 0, \ldots, 2^k - 1$, such that $\bigcup_{Q_l} = \mathcal{U}$. Assume that a one-to-one predefined mapping $p = \beta(l)$ exists between each possible k-bit information sequences $\mathbf{b}_l$, $l = 0, \ldots, 2^k - 1$ and the subsets Q_p, $p = 0, \ldots, 2^k - 1$. This means that each k-bit information sequence can be associated to one out of 2^{w-k} carrier codewords $\mathbf{u}_i$. Of course we must define a strategy to solve the above ambiguity, i.e. we must define how the carrier codeword is chosen among all the codewords in the same bin. Let us start by considering that this strategy has already been defined, and let us indicate the chosen carrier codeword by $\mathbf{u}_m$. We will go back to the choice of $\mathbf{u}_m$ at the end of this section.

We now consider the case in which an AWGN attack is present. In this scenario, denoting by $\mathbf{c}_w$ the transmitted n-dimensional column vector, the received n-dimensional column vector $\mathbf{r}$ can be expressed as:

$$\mathbf{r} = \mathbf{c}_w + \mathbf{n}, \tag{1}$$

$\mathbf{n}$ being an additive white Gaussian noise vector with variance σ_n^2, i.e., $\mathbf{n} \sim N(0, \sigma_n)$.

Upon receiving a sequence $\mathbf{r}$, the decoder performs the estimation of the $\hat{i}$-th carrier sequence by evaluating:

$$\hat{i} = \underset{i=0,\ldots,n-1}{\arg\max} \left(\mathbf{r}^T \mathbf{u}_i\right) \tag{2}$$

[2] The set with the n columns of $\mathbf{U}$ gives the codebook $\mathcal{U}$

where T stands for transpose operation. The estimated transmitted sequence $\mathbf{b}_{\hat{l}}$ corresponds to the sequence associated to the bin $\mathbf{u}_{\hat{i}}$ belongs to. Note that the decoding rule outlined above, together with the equi-energetic nature of the carrier codewords, ensures that the watermark is robust against multiplication by a scale factor g.

2.1 Constant Robustness Embedding

In order to derive the optimum fixed-robustness embedding strategy, a parameter measuring the robustness of the watermark is needed. To this aim, we propose to use the maximum pairwise error probability between the transmitted codewords and all the codewords of $\mathcal{U}$ belonging to a bin Q_j with $j \neq l$, where by l we indicated the index associated to the transmitted information sequence. Even if such a probability does not coincide with the true error probability of the system, it can be shown [13] that if the attack noise is not too strong, the maximum pairwise error probability is a good approximation of the true error probability[3].

With the above observations in mind, and by denoting with $P_e(m,q)$ the pairwise (error) probability that the receiver decodes the sequence $\mathbf{u}_q$ instead of the carrier sequence $\mathbf{u}_m$, we have:

$$P_e(m,q) = \text{Prob}\left\{ {\mathbf{c}_w}^T (\mathbf{u}_m - \mathbf{u}_q) + \mathbf{z} < 0 \right\} \tag{3}$$

where $\mathbf{z} \sim N\left(0, \sigma_n \sqrt{|\mathbf{u}_m - \mathbf{u}_q|}\right)$. By exploiting the well known approximation [13]:

$$P_e(m,q) \cong \frac{1}{2} \exp\left\{ \left[-\frac{{\mathbf{c}_w}^T (\mathbf{u}_m - \mathbf{u}_q)}{\sqrt{2}\sigma_n \sqrt{|\mathbf{u}_m - \mathbf{u}_q|}} \right]^2 \right\}, \tag{4}$$

and by proceeding as in [10], the fixed robustness embedding problem can be formulated as follows: evaluate the transmitted n-dimensional column vector $\mathbf{c}_w$ that minimizes the distortion $\Delta = (\mathbf{c}_w - \mathbf{c})^T (\mathbf{c}_w - \mathbf{c})$, subject to the linear constraint:

$${\mathbf{c}_w}^T \mathbf{u}_m - {\mathbf{c}_w}^T \mathbf{u}_q \geq S \ , \ \forall q | \mathbf{u}_q \notin Q_l, \tag{5}$$

with:

$$S = 2\sqrt{P_c \times \left(10^{-\frac{\text{DNR}}{10}}\right) \times \log\left(\frac{1}{2P_e^*}\right)}, \tag{6}$$

where is is assumed that the attacker uses the maximum noise power allowed to him, P_e^* indicates the target error probability, $P_c = E[\|\mathbf{c}\|^2]$, and where DNR indicates the Data to Noise Ratio defined as

$$\text{DNR} = 10\log_{10}\left(\frac{P_c}{\sigma_n^2}\right). \tag{7}$$

[3] On the other hand, when the attack noise gets high, the system performance deteriorates rapidly, hence making the above analysis useless.

Since the columns of the unitary matrix $\mathbf{U}$ represent an orthonormal basis for $\mathbb{R}^n$, it is possible to express the error vector $\mathbf{e} = \mathbf{c}_w - \mathbf{c}$ as a linear combination of $\mathbf{u}_i$'s, i.e.,

$$\mathbf{e} = \mathbf{U}\mathbf{a}, \tag{8}$$

where $\mathbf{a} = (a_0, a_1, \ldots, a_{n-1})^T$ is the column vector with the weights of the linear combination. Given the above, it is straightforward to observe that $\Delta = \|\mathbf{a}\|^2 = \sum_{h=0}^{n-1} a_h^2$ and $\mathbf{a}^T\mathbf{U}^T\mathbf{u}_i = a_i$. Accordingly, our problem is equivalent to find the vector $\mathbf{a}$ such that:

$$\begin{gathered} \mathbf{a} = \arg\min_{a_h} \left(\sum_{h=0}^{n-1} a_h^2 \right) \\ subject\ to: \\ a_m - a_q + \mathbf{c}^T\mathbf{u}_m - \mathbf{c}^T\mathbf{u}_q \geq S \ \ , \ \ q|\mathbf{u}_q \notin Q_l \end{gathered} \tag{9}$$

or:

$$\begin{gathered} \mathbf{a} = \arg\min_{a_m,\, a_q} \left(a_m^2 + \sum_{q|\mathbf{u}_q \notin Q_l} a_q^2 \right) \\ subject\ to: \\ a_q \leq a_m - S + \chi_{q,m} \ \ , \ \ q|\mathbf{u}_q \notin Q_l \end{gathered} \tag{10}$$

where $\chi_{q,m} = \mathbf{c}^T\mathbf{u}_m - \mathbf{c}^T\mathbf{u}_q$. The constraint in (10) can be reformulated as:

$$a_q = \min\left(0, a_m - S + \chi_{q,m}\right) \tag{11}$$

Indeed, if $a_m - S + \chi_{q,m}$ is greater than or equal to zero, the value of a_q which minimizes the error Δ while fulfilling the constraint is $a_q = 0$. Conversely, if $a_m - S + \chi_{q,m}$ is lower than zero the minimum is obtained at the edge, i.e., for $a_q = a_m - S + \chi_{q,m}$. Accordingly, the minimization problem can be expressed as:

$$\begin{gathered} \mathbf{a} = \arg\min_{a_m} \left(a_m^2 + \sum_{q \in C_l} a_q^2 \right) \\ a_q = \min\left(0, a_m - S + \chi_{q,m}\right)\ q|\mathbf{u}_q \notin Q_l \end{gathered} \tag{12}$$

Note that the problem is now formulated as a mono dimensional minimization problem in the unknown a_m. Such a minimum can be easily computed by means of a numeric approach (e.g., see [14]).

Having defined the optimum embedding rule, we now go back to the choice of $\mathbf{u}_m$. By recalling that the decoder takes its decision by maximizing the correlation between $\mathbf{r}$ and all the codewords in $\mathcal{U}$, we decided to choose the carrier codeword which maximizes the correlation with $\mathbf{c}$, i.e.

$$\mathbf{u}_m = \arg\max_{\mathbf{u}_s \in Q_l} \mathbf{c}^T\mathbf{u}_s \ . \tag{13}$$

3 Fixed Distortion Embedding

We now want to re-formulate the embedding problem by fixing the distortion Δ and maxime the watermark robustness, i.e. minimize the maximum pairwise error probability. To do so, we can rely on the analysis reported in the previous section, however it is necessary that a closed form expression for Δ is obtained. Let us start by denoting with $\tilde{\chi}_{q,m}$ the reordered set of $\chi_{q,m}$, so that $\tilde{\chi}_{0,m} \leq \tilde{\chi}_{1,m}, \ldots, \leq \tilde{\chi}_{d-1,m}$, where d is the dimension of $\chi_{q,m}$, i.e., $d = n - 2^{w-k}$. Of course, the unknown term a_m will satisfy one of the following mutually exclusive conditions:

$$\begin{aligned} &\text{(I)} && S - a_m < \tilde{\chi}_{0,m} \\ &\text{(II)} && \tilde{\chi}_{0,m} \leq S - a_m < \tilde{\chi}_{d-1,m} \\ &\text{(III)} && S - a_m \geq \tilde{\chi}_{d-1,m} \end{aligned} \tag{14}$$

Let us first assume that condition (I) holds. In this case, since $\tilde{\chi}_{0,m} \leq \tilde{\chi}_{q,m}$, it is also verified $a_m - S + \tilde{\chi}_{q,m} \geq 0$, that is, directly from (12), $a_q = 0$. Besides, since in this case the minimization function is $\Delta_m = a_m^2$ and since for hypothesis $a_m > S - \tilde{\chi}_{0,m}$, we have

$$a_m^{(0)} = \max\left(0, S - \tilde{\chi}_{0,m}\right), \tag{15}$$

and

$$\Delta_m^{(0)} = \left[\max\left(0, S - \tilde{\chi}_{0,m}\right)\right]^2, \tag{16}$$

where the apex 0 means that a_m and Δ_m are evaluated by assuming $S - a_m < \tilde{\chi}_{0,m}$.

If case (II) holds, it is of course possible to find an index f, for which:

$$\tilde{\chi}_{f,m} \leq S - a_m < \tilde{\chi}_{f+1,m}. \tag{17}$$

Hence, $a_m - S + \tilde{\chi}_{q,m} > 0$, for $q > f$, and $a_m - S + \tilde{\chi}_{q,m} \leq 0$, for $q \leq f$. We thus obtain directly from (12) $a_q = 0$, for $q > f$, and $a_q = a_m - S + \tilde{\chi}_{q,m}$, for $q \leq f$. The distortion becomes:

$$\Delta_m = a_m^2 + \sum_{q=0}^{f} \left(a_m - S + \tilde{\chi}_{q,m}\right)^2. \tag{18}$$

Since in this case Δ_m is a quadratic form of a_m, the computation of the minimum distortion subject to (17), say it $\Delta_m^{(f)}$, is straightforward. Indeed, since the derivative of (18) is zero for

$$a_m = \hat{a}_m = \frac{S(f+1)}{f+2} - \frac{\sum_{q=0}^{f} \tilde{\chi}_{q,m}}{f+2}, \tag{19}$$

the value of a_m which gives the minimum, call it $a_m^{(f)}$, is:

$$a_m^{(f)} = \begin{cases} \hat{a}_m, & \text{for } S - \tilde{\chi}_{f+1,m} \leq \hat{a}_m < S - \tilde{\chi}_{f,m} \\ S - \tilde{\chi}_{f+1,m}, & \text{for } \hat{a}_m < S - \tilde{\chi}_{f+1,m} \\ S - \tilde{\chi}_{f,m}, & \text{for } \hat{a}_m \geq S - \tilde{\chi}_{f,m}. \end{cases} \tag{20}$$

Note that for high values of f, equation (19) can be rewritten as:

$$a_m^{(f)} \cong S - \frac{\sum_{q=0}^{f} \tilde{\chi}_{q,m}}{f+1} \geq S - \tilde{\chi}_{f,m}. \tag{21}$$

We will assume in the following that (21) holds for each f. By considering (21) and (20), we obtain:

$$\begin{aligned} a_m^{(f)} &= S - \tilde{\chi}_{f,m}, \\ \Delta_m^{(f)} &= \left(S - \tilde{\chi}_{f,m}\right)^2 + \sum_{q=0}^{f} \left(\tilde{\chi}_{q,m} - \tilde{\chi}_{f,m}\right)^2. \end{aligned} \tag{22}$$

Finally, by means of similar considerations, we have for case (III):

$$\begin{aligned} a_m^{(d-1)} &= S - \tilde{\chi}_{d-1,m}, \\ \Delta_m^{(d-1)} &= \left(S - \tilde{\chi}_{d-1,m}\right)^2 + \sum_{q=0}^{d-1} \left(\tilde{\chi}_{q,m} - \tilde{\chi}_{d-1,m}\right)^2. \end{aligned} \tag{23}$$

According to the above considerations, the distortion minimization problem can be expressed as:

$$\begin{aligned} h_m &= \underset{h=0,\ldots,d-1}{\arg\min}\ \Delta_m^{(h)} \\ a_m &= a_m^{(h_m)} \\ \Delta_m &= \Delta_m^{(h_m)}. \end{aligned} \tag{24}$$

Note that (24) allows to evaluate the minimum distortion for a given robustness S and a given m. Such an estimation can be performed by computing all the d possible values of the error Δ_m and selecting the minimum.

The above procedure can be easily managed so that the inverse problem, that is to evaluate the maximum robustness S for a given error Δ, is addressed. Firstly, observe from (22) that a given error Δ can be achieved only if

$$\sum_{q=0}^{f} \left(\tilde{\chi}_{q,m} - \tilde{\chi}_{f,m}\right)^2 < \Delta \tag{25}$$

Accordingly, the search must be restricted to the set of values f which satisfy (25), say $\{0, 1, \ldots, d'-1\}$, with $d' \leq d$. Now, for a given Δ, it is possible to derive from (23) and (22) the robustness parameter $S_m^{(h)}$, with $h \in \{0, 1, \ldots, d'-1\}$, as:

$$S_m^{(h)} = \tilde{\chi}_{h,m} + \sqrt{\Delta - \sum_{q=0}^{h} \left(\tilde{\chi}_{q,m} - \tilde{\chi}_{h,m}\right)^2}, \tag{26}$$

Accordingly, the maximum robustness problem can be expressed as:

$$\begin{aligned} p_m &= \underset{p=0,\ldots,d'-1}{\arg\max}\ S_m^{(p)} \\ a_m &= a_m^{(p_m)}, \end{aligned} \tag{27}$$

Note that both (27) and (24) can be evaluated by means of an exhaustive procedure over all d possible values of $\Delta_m^{(h)}$ and $S_m^{(h)}$, respectively.

3.1 Quasi-Orthogonal Dirty Paper Coding

As in [10], to further improve the performance of the proposed system we replace the orthogonal codes with quasi-orthogonal sequences, so to increase the number of available codewords for a given sequence length n. Specifically, we use Gold sequences of length n since their cross-correlation properties ensure that different sequences are almost orthogonal among them [13]. Accordingly, the matrix $\mathbf{U}$ is now a rectangular $n \times h$ matrix with column vectors $\mathbf{u}_i$, $i = 1, \ldots, h$, representing a set of h Gold sequences with length n. Gold sequences have been widely studied in the technical literature, particularly for spread spectrum applications, for their autocorrelation and cross-correlation functions that are reminiscent of the properties of white noise. Specifically, in the following we will assume that $\mathbf{u}_i$ are normalized Gold sequences [15] with $u_i(l) = \pm\frac{1}{\sqrt{n}}$, $\forall i, l$. Note that all Gold sequences have the same norm, thus ensuring that the decoder performance are invariant with respect to multiplication by a gain factor g. In this case, for a given length $n = 2^w - 1$, the number of possible Gold sequences that are characterized by good periodic cross-correlation properties is $n + 2$. Since each cyclic shift of any Gold sequence is still characterized by the same properties, the overall number of Gold sequences that can be considered for information embedding is $h = n(n + 2)$. Note that, as required to write (8), Gold sequences are a frame for $\mathbb{R}^n$, hence ensuring that every element of $\mathbb{R}^n$ can be expressed as a linear combination of the $\mathbf{u}_i$'s.

Let us now consider the distortion introduced by watermark embedding, we have:

$$d = \left| \sum_{i=1}^{h} a_i \mathbf{u}_i \right|^2 = \sum_{i=1}^{h} a_i^2 + \sum_{i \neq j} a_i a_j \mathbf{u}_i^T \mathbf{u}_j. \tag{28}$$

We can argue that, due to the particular properties of Gold sequences, the first term of the above equation is predominant with respect to the second one, even if the second term is not exactly equal to zero due to the non perfect orthogonality of Gold sequences. Such an assumption will be validated through numerical simulations in section 6. By relying on the above observations, the fixed distortion constraint can still be replaced by a constraint on $\sum_{i=1}^{m} a_i^2$.

4 Perceptual Dirty Paper Coding

The analysis carried out in the previous section gives the possibility of fixing the embedding distortion. This turns out to be a very useful feature if we want to give to the embedding systems a perceptually-flavored behavior. More specifically, we consider the watermarking of still images in the block-DCT domain. The host image is first partitioned into non-overlapping 8×8 blocks, that then are DCT-transformed. For each DCT block a set of intermediate frequency coefficients is extracted to form the host feature vector. In our implementation we considered 12 DCT coefficients for each block, more specifically after zig-zag scanning the DCT block we skip the first 3 coefficients and select the next 12 ones.

At this point we need a system to measure the maximum amount of distortion that can be tolerated by each coefficient before the watermark becomes visible. Though

many algorithms are available to this aim, we decided to adopt the approach proposed by Watson in [11] for its simplicity.

At a general level Watson's visual model consists of three main parts: a sensitivity function giving the visibility of a visual stimulus as a function of frequency; two masking components, taking into account the capacity of the host image to mask the stimulus; and a pooling function to consider how visual stimuli at different frequencies combine together to form the final visual appearance of the composite stimulus.

The sensitivity function is given as a table specifying for each DCT position the smallest magnitude (Just Noticeable Difference - JND) of the corresponding DCT coefficient that is visible in the absence of any masking components. Let us denote the, so to say, threshold values contained in the sensitivity table by $t(i,j)$, where the indexes i and j indicate the position of the DCT coefficient within the 8×8 block. The exact values of the sensitivity table depends on a number of parameters, including viewing conditions, environment lightness, etc. Here we used the values given in [12]. To take into account luminance masking, Watson suggests to modify the threshold values as:

$$t_l(i,j,k) = t(i,j) \left(\frac{C(0,0,k)}{C_{0,0}} \right)^{0.649}, \tag{29}$$

where $C(0,0,k)$ is the DCT coefficient of the k-th block and $C_{0,0}$ is the average value of all the DCT coefficients of the image. Note that the modified thresholds vary from block to block due to the presence of the $C(0,0,k)$ term. Finally, the modified thresholds $t_l(i,j,k)$ are adjusted to take into account iso-frequency contrast masking, leading to a final masked threshold (or slack) given by:

$$s(i,j,k) = \max\{t_l(i,j,k); \|C(i,j,k)\|^{0.7} t_l(i,j,k)^{0.3}\}. \tag{30}$$

Of course a different $s(i,j,k)$ is obtained for each coefficient, however in our case we need to specify the same distortion, for all the n coefficients bearing the same bit. For this reason the embedder considers an average distortion computed as:

$$\Delta^2_{max,av} = \frac{\sum s(i,j,k)^2}{n}, \tag{31}$$

where the sum is extended to all the n coefficients hosting the same bit. Note that since typically n is larger than 12, the sum spans several DCT blocks. For instance, for $n = 32$, the sum spans three blocks[4].

At this point the fixed distortion embedding algorithm described in the previous section is applied to embed the bit of the information message into the host features. Note that a different distortion constraint is applied to DCT blocks hosting different bits, hence each bit will be characterized by a different robustness.

5 Multistage Decoding

As we pointed out at the end of the previous section, bit hosted by different DCT blocks are characterized by different levels of robustness. As an extreme case, for some blocks

[4] We neglect border effects for simplicity.

the admissible distortion may be so low that the embedding algorithm fails to enter the correct decoding region. In other words, in certain regions the interference of the host image can not be rejected completely, leading to a non-null error probability even in the absence of attacks. In order to improve the robustness of the watermark, an additional channel coding step prior to orthogonal (or Gold) dirty paper coding is introduced. More specifically a turbo coding (decoding) step is performed prior to watermark embedding. To this aim, let us observe that the detection strategy (2) generates hard estimates of the bits $\mathbf{b}_l = (b_{l,0}, \ldots, b_{l,k-1})$. On the other hand, when dealing with multistage decoding it is preferable that the inner decoder produces soft estimates to be delivered to the outer decoder [13]. In order to provide the outer decoder with a soft estimate of the hidden bit, we follow the same approach described in [10]. Let the sets $I_{1,s}$ and $I_{0,s}$ be defined as:

$$\begin{aligned} I_{1,s} &= \{l : b_{l,s} = 1\}, \\ I_{0,s} &= \{l : b_{l,s} = 0\}, \end{aligned} \tag{32}$$

that is $I_{1,s}$ ($I_{0,s}$) represents the set of 2^{k-1} sequences $\mathbf{b}_l$ for which the s-th bit is 1 (0). Then we use the following soft estimate of the s-th bit:

$$v_s = P_{1,s} - P_{0,s} = \max_{\mathbf{u}_i \in Q_l, l \in I_{1,s}} \left(\mathbf{r}^T \mathbf{u}_i\right) - \max_{\mathbf{u}_i \in Q_l, l \in I_{0,s}} \left(\mathbf{r}^T \mathbf{u}_i\right). \tag{33}$$

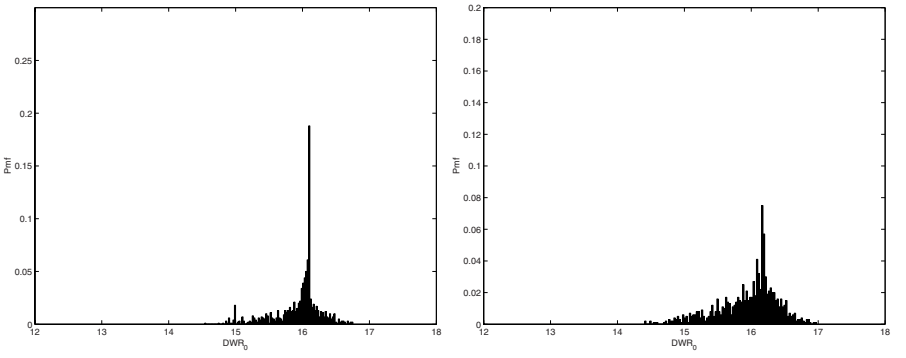
The sign of (33) determines the hard estimate of the s-th bit and its absolute value represents the soft output information that can be used by the outer decoder.

It is worth pointing out that the above soft decoding strategy can be applied to any kind of binary outer coder's structure. In this paper, the outer code is the $R_c = 1/2$ binary punctured parallel concatenated turbo coder presented in [16] which allows to achieve error correction performance that are very close to the theoretical Shannon limit.

We conclude this section by highlighting the necessity of scrambling the to-be-hidden bits after the turbo encoder, prior to embedding. This is because due to the coherence of natural still images, the DCT blocks characterized by a very low admissible distortion are likely to be contiguous, hence resulting in the introduction of bursty errors. The scrambler avoids this problem by transforming bursty errors into isolated errors. Of course, de-scrambling is applied at the decoder prior to turbo decoding.

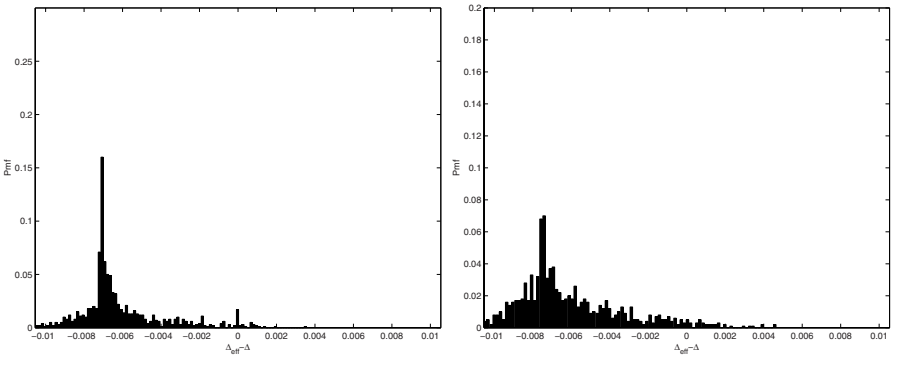
6 Simulations and Experimental Results

The validity of the above analysis and the performance of the watermarking scheme deriving from it, have been tested by means of both numerical simulations and experimental tests. Simulations aimed at validating the fixed distortion embedding strategy derived theoretically. This is a necessary step when we use Gold sequences instead of orthogonal codewords, since the analysis we carried out relies on the assumption that the second term in equation (28) is negligible with respect to the first one. In figure 1 the histogram of the actual embedding distortion d (measured in terms of of DWR) when a target DWR of 15dB was asked is shown. The histogram was built by applying the



(a) (b)

Fig. 1. Histogram of the actual DWR when a target DWR of 15dB is asked, for Gold sequences of length 32 (a) and 64 (b).



(a) (b)

Fig. 2. Histogram of the second term in equation (28) when a target DWR of 15dB is asked, for Gold sequences of length 31 (a) and 63 (b). The histograms should be compared with the value of $\sum_i a_i^2$, that, for DWR = 15dB, is approximately equal to 0.0316 (we let $P_c = 1$).

embedding algorithm to several cover sequences. As it can be seen the actual DWR is slightly higher than the target one. In figure 2 the histogram of the second term in equation (28) is plotted (linear scale). As it can be verified the error we made by neglecting this term is negligible. As a matter of fact with DWR = 15dB, and since in our simulations we let $P_c = 1$, we have that $\sum_i a_i^2 = 10^{-1.5} \approx 0.0316$ which is much higher than the values reported in figure 2. In addition in most of the cases this term turns out to be negative, hence ensuring that the actual distortion is lower than the target one.

In order to estimate the overall performance of the system, a selection of the results we obtained on real images is now described. For sake of brevity we describe only the performance of the algorithm based on Gold sequences. Similar results (actually,

slightly worse) were obtained for the orthogonal case, which, on the other hand, ensured a much faster embedding phase.

6.1 Watermark Invisibility

We fist checked whether the proposed fixed distortion strategy actually ensures the invisibility of the watermark. To do so, we built a database of 40 1024 $\times$ 1024 images, and embedded the watermark in all of them by letting $n = 32$ and $k = 1, 2$, thus obtaining an overall rate of 1/64 and 1/32 respectively. We visually inspected all the marked images and the watermark resulted to be invisible in all the cases: the observer could individuate the watermark only by comparing two magnified versions of the original and watermarked images on a high resolution monitor. No visual artifact was perceived by looking at the images in normal conditions or by looking at the images after printing by means of a high quality printer.

For all the images we measured the DWR (data to watermark ratio) both by considering only the watermarked DCT coefficients and globally, i.e. by exploiting the fact the not all the DCT coefficients are marked. The results we obtained are reported in table 1. In the same table, the Watson distance [12] between the watermarked and the original images is also given.

Table 1. Objective measures of the distortion introduced by the watermark. The results have been obtained by averaging those obtained on a test database consisting of 40 1024 $\times$ images. By DWR$_{all}$, DWR$_{sel}$ and D$_{wats}$, the DWR computed on the overall image, the host DCT coefficients and the Watson distance are meant respectively.

Rate	DWR$_{all}(db)$	DWR$_{sel}$ (db)	D$_{Wats}$ (db)
$n = 32$, $k = 2$	37.46	13.24	17.43
$n = 32$, $k = 1$	37.52	13.12	17.49

6.2 Watermark Robustness

With regard to robustness, given the fixed distortion embedding strategy we adopted, we first had to evaluate whether host signal rejection was actually achieved or not (the admitted distortion could not be enough to ensure that the right decoding region is entered). Hence we tried to detect the watermark on the marked images in the absence of attacks. We repeated this test on all the images of the database and no errors were found. Then we considered a number of attacks involving scaling (not necessarily uniform) of the host features. In particular we considered histogram stretching, histogram equalization and sharpening. In all the cases the watermark was successfully recovered with no errors in all the images of the database. To give an idea of the robustness of our system against this kind of attacks, two examples of images attacked by means of histogram equalization are shown in figure 3. As it can be seen the attack strength may be very high, and amplitude scaling of DCT coefficients highly non-uniform, nevertheless the watermark is correctly retrieved.

As a second test we considered robustness against white noise addition. More specifically the watermarked image was impaired by spatially adding a white Gaussian noise, with increasing variance. The results we obtained demonstrate only a moderate robustness against this kind of attack. For example, when the variance of noise is set to 10, the bit error probability was equal to $1.1 \cdot 10^{-1}$ ($k = 1$). Note that adding a white Gaussian noise with variance 10 results in a visible, yet slight, degradation of the marked image. It has to be noted, though, that such an attack results in an average WNR - computed only on the host features - approximately equal to -2 db, and that for negative WNR values, a high robustness can only be achieved for lower rates (or by relaxing the invisibility constraint).

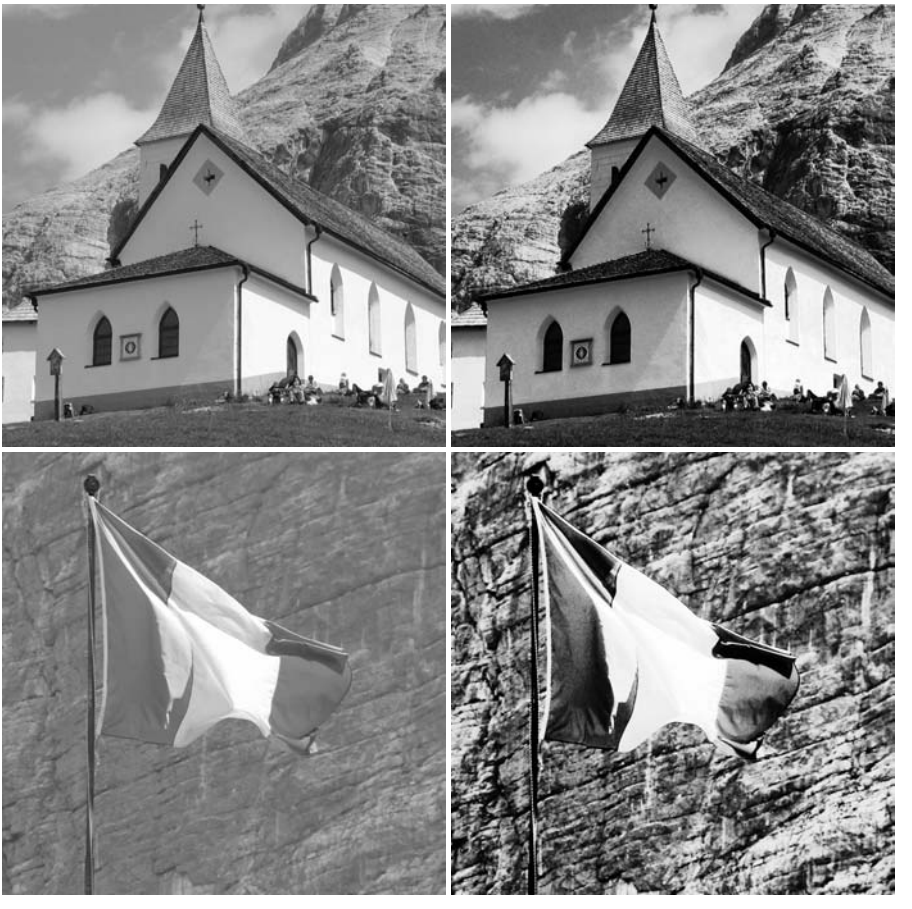


Fig. 3. Robustness against histogram equalization. Despite the great difference between the marked (left) and the marked and attacked (right) images, no decoding error was found.

Similar considerations hold when robustness against JPEG compression is considered. The results we obtained in this case are summarized in table 2.

Table 2. Robustness against JPEG compression. The bit error probability averaged over all the images of the database is given as a function of the JPEG quality factor (Q).

Rate	$Q = 90$	$Q = 80$	$Q = 70$
$n = 32, k = 2$	0	$1.2 \cdot 10^{-2}$	0.34
$n = 32, k = 1$	0	$3 \cdot 10^{-3}$	$1.2 \cdot 10^{-1}$

7 Conclusions

By relying on the simple structure of orthogonal and Gold sequences, we have presented a new dirty paper coding watermarking scheme. The main merit of the proposed scheme is the use of an optimum embedding strategy, which permits to maximize the robustness of the watermark for a fixed distortion. Another advantage of the new scheme is that due to the equi-energetic nature of the codewords and to the adoption of a correlation-based decoder, robustness against value-metric scaling is automatically achieved, thus achieving a very good robustness against common image processing tools such as image enhancement and histogram manipulation. We have also shown how the performance of the system are improved by concatenating the dirty paper code with an outer turbo code. To this aim, we had to introduce a new soft dirty paper decoding scheme which allows the iterative multistage decoding of the concatenated codes. The validity of the proposed techniques has been assessed through experimental results which demonstrated an excellent behaviour from the point of view of watermark invisibility and robustness against attacks involving scaling of the host features.

Several directions for future work remain open, including the usage of more powerful spherical codes [17, 18, 19] instead of the simple orthogonal codes used here and the adoption of more sophisticated HVS models to improve watermark invisibility.

References

[1] Eggers, J.J., Girod, B.: Informed Watermarking. Kluwer Academic Publishers (2002)
[2] El Gamal, A., Cover, T.M.: Multiple user information theory. Proceedings of the IEEE **68** (1980) 1466–1485
[3] Cover, T.M., Thomas, J.A.: Elements of Information Theory. Wiley, New York (1991)
[4] Chen, B., Wornell, G.: Quantization index modulation: a class of provably good methods for digital watermarking and information embedding. IEEE Trans. on Information Theory **47** (2001) 1423–1443
[5] Eggers, J.J., Bäuml, R., Tzschoppe, R., Girod, B.: Scalar Costa scheme for information embedding. IEEE Trans. on Signal Processing **4** (2003)
[6] Perez-Gonzalez, F., Balado, F., Hernandez, J.R.: Performance analysis of existing and new methods for data hiding with known-host information in additive channels. IEEE Trans. on Signal Processing **51** (2003) 960–980

[7] Chou, J., Ramchandran, K.: Robust turbo-based data hiding for image and video sources. In: Proc. 9th IEEE Int. Conf. on Image Processing, ICIP'02. Volume 2., Rochester, NY, USA (2002) 133–136
[8] M. L. Miller, G. J. Doerr, I.J.C.: Dirty-paper trellis codes for watermarking. In: Proc. 9th IEEE Int. Conf. on Image Processing, ICIP'02. Volume II., Rochester, NY (2002) 129–132
[9] Miller, M.L., Doerr, G.J., Cox, I.J.: Applying informed coding and embedding to design a robust, high capacity, watermark. IEEE Trans. on Image Processing **13** (2004) 792–807
[10] Abrardo, A., Barni, M.: Orthogonal dirty paper coding for informed watermarking. In Wong, P.W., Delp, E.J., eds.: Security, Steganography, and Watermarking of Multimedia Contents VI, Proc. SPIE Vol. 5306, San Jose, CA, USA (2004)
[11] Watson, A.B.: DCT quantization matrices visually optimized for individual images. In Allebach, J.P., Rogowitz, B.E., eds.: Human Vision, Visual Processing, and Digital Display, Proc. SPIE vol. 1913, San Jose, CA (1993) 202–216
[12] Cox, I.J., Miller, M.L., Bloom, J.A.: Digital Watermarking. Morgan Kaufmann (2001)
[13] Proakis, J.G.: Digital Communications, 2nd Edition. McGraw-Hill, New York (1989)
[14] Forsythe, G.E., Malcolm, M.A., Moler, C.B.: Computer Methods for Mathematical Computations. Prentice Hall (1976)
[15] Gold, R.: Optimal binary sequences for spread spectrum multiplexing. IEEE Trans. on Information Theory **13** (1967) 619–621
[16] Berrou, C., Glavieux, A., Thitimajshima, P.: Near shannon limit error-correcting coding and decoding: Turbo-codes. In: Proceedings of ICC, IEEE International Conference on Communications, Geneva, Switzerland (1993) 1064–1070
[17] Conway, J.H., Sloane, N.J.A.: Sphere Packings, Lattices, and Groups. Springer-Verlag, New York (1988)
[18] Hamkins, J., Zeger, K.: Asymptotically dense spherical codes - part I: wrapped spherical codes. IEEE Trans. on Information Theory **43** (1997) 1774–1785
[19] Hamkins, J., Zeger, K.: Asymptotically dense spherical codes - part II: laminated spherical codes. IEEE Trans. on Information Theory **43** (1997) 1786–1798