

# THE IMPACT OF CHANNEL CODING ON THE PERFORMANCE OF SPATIAL WATERMARKING FOR COPYRIGHT PROTECTION

*J. R. Hernández, F. Pérez-González and J. M. Rodríguez*

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

## ABSTRACT

In this paper we analyze the effect that the application of channel coding produces on the performance of the watermark detection and decoding tests for copyright protection of images. Detector structures are derived for both tests and analytical bounds and approximations are obtained for the bit error rate (BER) and the receiver operating characteristic (ROC) associated with the watermark decoding and detection tests when block codes are employed. The extension to other families of codes is discussed. Finally, the analytical expressions are contrasted with experimental results in several cases of interest.

## 1. INTRODUCTION

The enormous progress that digital technologies have experienced during the last decades has contributed to popularize the use of electronic media for transmission and storage of documents, images, audio, video and other types of information. Information stored in digital format can be copied without quality loss and distributed efficiently at fairly low cost. These developments have also increased the potential for interception, manipulation, misuse and unauthorized distribution of information. This is in fact one of the main impediments to commercial use of communication networks and electronic storage media for distribution of digital information. For this reason, the design of techniques for preserving the ownership of digital information is fundamental to the development of future multimedia services.

Previous research on copyright protection of still images has resulted in the appearance of several methods based on watermarking. In all these techniques the contents of the original image are altered in a fashion determined by a secret key and, optionally, by a certain amount of information to be hidden into the image. Some of these methods perform the watermarking process in the spatial domain using spread spectrum techniques [7, 3]. Other methods add the watermark in the frequency domain by computing the DCT of the whole image [2] or in a block basis [8, 7, 1, 3].

Even though different proposals for solving the copyright enforcement problem have been described and tested with diverse results, previous research in watermarking techniques has suffered from the absence of a theoretical approach to the limits in performance of these methods. In this paper we study how the introduction of channel coding affects the performance of a watermarking system based on a 2D-multipulse modulation. Section 2 presents this modulation and the equivalent Gaussian vector channel that results. In sections 3 and 4 we analyze the performance of the watermark decoding and detection process with channel codes. Finally, in section 5 we present results from simulations.

## 2. 2-D MULTI-PULSE AMPLITUDE MODULATION

### 2.1. Definitions

In a 2D-multipulse amplitude modulation watermarking scheme [4, 5], the watermark can be expressed as a linear combination of  $L$  orthogonal functions  $\{p_i[m, n]\}$ ,  $i \in \{1, \dots, L\}$ :

$$w[m, n] = \sum_{i=1}^L b_i p_i[m, n], \quad (1)$$

where the coefficients  $b_1, \dots, b_L$  encode a hidden message. The watermark  $w[m, n]$  is added to the original image  $x[m, n]$  to obtain the watermarked version  $y[m, n] = x[m, n] + w[m, n]$ . In the scheme we are considering [4, 5], the pulses  $p_i[m, n]$  are defined as

$$p_i[m, n] \triangleq \begin{cases} \alpha[m, n] s[m, n] & \text{if } (m, n) \in \mathcal{S}_i \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

where both  $s[m, n]$  and the sets  $\{\mathcal{S}_i\}$ , are generated as a function of a secret key  $K$  to provide cryptographic security. The signal  $s[m, n]$  is the output of a pseudonoise generator, and is modeled, considering  $K$  as a random variable, as an uncorrelated random sequence. We propose the use of key-dependent sparse pulses spread out over the whole image to add spatial uncertainty about the locations where hidden bits are placed and to increase the resilience to cropping. We will assume in the sequel non-overlapping pulses, i.e.  $\mathcal{S}_i \cap \mathcal{S}_j = \emptyset \quad \forall i \neq j$ , so the pulses will always be orthogonal. This modulation technique is similar to a direct-sequence spread spectrum (SS) scheme. However, the main difference with respect to classical SS systems is that in our context the jammer is not limited to additive noise attacks. He can in fact play the role of a worst-case channel especially designed to attack the hidden signal without perceptually degrading the image.

### 2.2. Equivalent Vector Channel

Due to the lack of good statistical models for images, we reduce the observation space to the projection of the image onto the pulses  $\{p_i\}$  and assume that the information in the subspace orthogonal to these pulses can be ignored. We assume that the original image  $x[m, n]$  is not available in the detection process and that the watermarked image is filtered with a space-variant linear filter  $h_{ki}[m, n]$ , which models a filtering attack or a preprocessing step before detection, obtaining a signal  $z[m, n]$  as a result. Let us define  $x^{k,l}[m, n] = x[m - k, n - l]$  and similarly  $y^{k,l}[m, n]$  and  $p_i^{k,l}[m, n]$ . Then,

$$r_i \triangleq \langle z, p_i \rangle = \sum_{j=1}^L b_j \sum_{k,l} \langle h_{k,l} p_j^{k,l}, p_i \rangle + \sum_{k,l} \langle h_{k,l} x^{k,l}, p_i \rangle \quad (3)$$

We assume that the probability of a pixel  $(m, n)$  being assigned to any set  $\mathcal{S}_i$  is  $1/L$  and that the assignment is done independently for each pixel. We will model the vector  $\mathbf{r} = (r_1, \dots, r_L)$  statistically for a fixed image  $x[m, n]$ , treating the key  $K$  as the only random variable in the model. The vector  $\mathbf{r}$  can be expressed as [4, 5]:

$$\mathbf{r} = \mathbf{A}\mathbf{b} + \mathbf{n} \quad (4)$$

where  $\mathbf{b} = (b_1, \dots, b_L)$ ,  $\mathbf{A}$  is a deterministic diagonal matrix and  $\mathbf{n}$  is a zero-mean uncorrelated Gaussian random vector. Let  $\mathbf{\Gamma}$  be the covariance matrix of  $\mathbf{n}$ . The elements of  $\mathbf{A}$  and  $\mathbf{\Gamma}$  are [4]:

$$a_{ij} = \delta_{ij} \frac{1}{L} \sum_{m,n} h_{0,0}[m, n] \alpha^2[m, n] \quad (5)$$

$$\begin{aligned} \gamma_{ii} &= \frac{1}{L} \sum_{m,n} \alpha^2[m, n] x_f^2[m, n] \\ &+ b_i^2 \frac{1}{L} \sum_{m,n} h_{0,0}^2[m, n] \alpha^4[m, n] (E[s^4] - 1) \\ &+ b_i^2 \frac{1}{L} \sum_{(k,l) \neq (0,0)} \sum_{m,n} h_{k,l}^2[m, n] \alpha^2[m, n] \alpha^2[m-k, n-l] \\ &+ b_i^2 \frac{L-1}{L^2} \sum_{m,n} h_{0,0}^2[m, n] \alpha^4[m, n] \end{aligned} \quad (6)$$

$$\gamma_{ij} = -b_i b_j \frac{1}{L^2} \sum_{m,n} h_{0,0}^2[m, n] \alpha^4[m, n], \quad i \neq j \quad (7)$$

where  $\delta_{ij}$  is the Kronecker delta function,  $x_f[m, n]$  is the image filtered by  $h_{kl}[m, n]$ . Even though  $\mathbf{\Gamma}$  is non-diagonal, the cross-covariance terms are small compared to the terms in the diagonal if  $L$  is large enough. Therefore,  $\mathbf{r}$  can be accurately modeled as the output of a memoryless Gaussian vector channel. We will assume in the sequel that  $b_i \in \{-1, 1\}$ ,  $\forall i$  and, as a consequence, that  $\gamma_{ii} = \gamma$ ,  $a_{ii} = a \forall i$ . The watermarked image could be attacked by adding zero-mean white noise. If the noise variance at pixel  $(m,n)$  is  $\sigma_n^2[m, n]$ , then we can analyze the effect of this attack just adding to Eq. (6) the term  $(\sum_{m,n} \alpha^2[m, n] \sigma_n^2[m, n])/L$ .

### 3. CHANNEL CODING

#### 3.1. Binary Antipodal Signaling

Suppose that codewords  $b_i(k) \in \{1, -1\}$   $i = 1, \dots, L, k = 1, \dots, M$  in a binary antipodal constellation are used to encode hidden messages. The bit-by-bit hard decoder is close to the optimal ML detector since the cross-covariance terms in the noise covariance matrix are negligible if  $L$  is large enough. The probability of bit error averaged over all the keys for a given image is:

$$P_b = Q\left(\frac{a}{\sqrt{\gamma}}\right) \quad (8)$$

Channel codes can be used to improve the performance of the data hiding system in terms of the bit error probability. From Eqs. (5) and (6) we infer that the SNR of the equivalent channel decreases as we increase the length of the encoded message. Hence, two are the main factors that determine the performance of a code when applied to the Gaussian channel derived in section 2.2: the minimum distance and the redundancy of the code. The best code for a given minimum distance is the one with minimum redundancy.

#### 3.2. Coding

If we use a bit-by-bit hard decoder, the result can be modeled as the output of a BSC (Binary Symmetric Channel) with parameter  $p = P_b$ , where  $P_b$  can be obtained from (8). The Bhattacharyya upper bound for the bit error probability of a  $(n, k)$  block code with minimum distance  $d_{min}$  when applied to this BSC is

$$P_b \leq \frac{M}{2(M-1)} \sum_{l=2}^M [4p(1-p)]^{w_l/2} \quad (9)$$

where  $M = 2^k$  is the number of codewords,  $w_l$  is the Hamming weight of the  $l^{\text{th}}$  codeword, and  $l = 1$  corresponds to the all-zeros codeword.

Similar bounds can be found for a convolutional code as a function of the parameter  $d_{free}$  that characterizes each code. The rate of the convolutional code also plays an important role, since the addition of redundancy produces a degradation of the SNR in the equivalent channel. The optimum ML detector is the minimum euclidean distance detector, since the channel is approximately Gaussian and memoryless. Therefore, a Viterbi algorithm implementation can be employed.

### 4. SYNC RECOVERY AND WATERMARK DETECTION

So far we have assumed that the exact location of the pulses was known. However, attacks such as cropping and affine transforms may change the spatial location of the watermark. The synchronization recovery algorithm and the watermark detection test are actually intimately related. When the former succeeds/fails to acquire synchronization, we can infer that the image is watermarked/not watermarked. Assume that the watermarked image  $z[m, n]$  may have suffered a geometric transformation  $T(\cdot)$  with unknown parameters  $\xi$  for which we do not suppose any a priori distribution. The watermark detection test can be formulated as the binary hypothesis test:

$$\begin{aligned} H_1 : z[m, n] &= T(x[m, n] + w[m, n], \xi) \\ H_0 : z[m, n] &= T(x[m, n], \xi) \end{aligned} \quad (10)$$

We will limit our analysis to transformations consisting in integer spatial shifts. Suppose also that  $L_s$  pulses are reserved for synchronization purposes and are thus modulated by known coefficients (assume +1). As we did in the decoding process, we will use the correlation coefficients  $r_i$  as the observations in the detection test. A uniformly most powerful (UMP) test does not exist in general. However, we can instead design the ML test assuming that  $\xi$  is correct and evaluate the resulting likelihood function at each possible  $\xi$ . If the likelihood function is greater than the threshold for some  $\xi$ , then we decide  $H_1$ . This procedure is equivalent to the test [4]:

$$l(z) = \max_{\xi} \frac{\sum_{i=1}^M f(\mathbf{r} | \mathbf{b}(i), \xi, H_1)}{f(\mathbf{r} | H_0, \xi)} \underset{H_0}{\overset{H_1}{\geq}} \eta \quad (11)$$

For every  $\xi$  the pdf  $f(\mathbf{r} | \mathbf{b}(i), \xi, H_1)$  can be approximated as a Gaussian pdf with mean  $\mathbf{A}\mathbf{b}(i)$  and covariance  $\mathbf{\Gamma}$  [4]. When the image is not watermarked,  $\mathbf{r} = \mathbf{n}$ , where  $\mathbf{n}$  is zero-mean white noise with variance

$$\gamma_0 = \frac{1}{L} \sum_{m,n} \alpha^2[m, n] x_f^2[m, n] \quad (12)$$

In the following sections we derive the likelihood tests conditioned to  $\xi$  for different channel coding schemes.

#### 4.1. Binary Antipodal Signaling

If we neglect the cross-covariance terms in  $\Gamma$ , we get the log maximum likelihood function:

$$l(z) = \max_{\xi} \frac{L}{2} \ln \frac{\gamma_0}{\gamma} - \frac{a^2 L}{2\gamma} + \frac{1}{2} \left( \frac{1}{\gamma_0} - \frac{1}{\gamma} \right) \sum_{i=1}^L r_i^2(\xi) + \frac{a}{\gamma} \sum_{i=1}^{L_s} b_i r_i(\xi) + \sum_{i=L_s+1}^L \ln \cosh \left( \frac{a r_i(\xi)}{\gamma} \right) \underset{H_0}{\overset{H_1}{\geq}} \eta$$

where  $r_i(\xi) = \langle z, T(p_i, \xi) \rangle$ .

#### 4.2. Coding

In this case the ML detector conditioned to a certain  $\xi$  is not practical due to its high computational complexity. Instead, we can define the following suboptimal test: first, obtain an estimate  $\hat{b}$  of the encoded message using a hard decisor and a minimum Hamming distance decoder; then, decide between the two hypothesis

$$\begin{aligned} H_1: & \quad r \text{ is watermarked with } \hat{b}. \\ H_0: & \quad r \text{ is not watermarked.} \end{aligned}$$

Then, the resulting watermark detection test is

$$l(z) = \max_{\xi} \frac{L}{2} \ln \frac{\gamma_0}{\gamma} - \frac{a^2 L}{2\gamma} + \frac{1}{2} \left( \frac{1}{\gamma_0} - \frac{1}{\gamma} \right) \sum_{i=1}^L r_i^2(\xi) + \frac{a}{\gamma} \sum_{i=1}^{L_s} b_i r_i(\xi) + \frac{a}{\gamma} \sum_{i=L_s+1}^L r_i(\xi) \hat{b}_i(\xi) \underset{H_0}{\overset{H_1}{\geq}} \eta \quad (13)$$

With convolutional codes the optimum ML detector is computationally complex. Therefore, the algorithm proposed for block codes is also applicable in this context. Hence,  $\hat{b}$  can be obtained using a Viterbi sequence detector and the final decision is made employing the function in (13).

#### 4.3. Performance Evaluation

Let  $\mu(s) = \ln E[e^{sI(\mathbf{r})}]$ . The probabilities of false alarm ( $P_F$ ) and detection ( $P_D$ ) can be bounded as follows [9]:

$$\begin{aligned} P_F &\leq e^{\mu(s) - s\hat{\mu}(s)} & s > 0 \\ P_D &\geq 1 - e^{\mu(s) + (1-s)\hat{\mu}(s)} & s < 1 \end{aligned} \quad (14)$$

and  $\hat{\mu}(s) = \eta$ . The maximization of the likelihood function can be difficult to implement when general affine transforms are possible, since the maximum is very narrow. Work is in progress for designing different pulse generation techniques which allow the use of combined brute force search and gradient optimization algorithms.

### 5. EXPERIMENTAL RESULTS AND COMPARISONS

In figure 1 we can see the 256x256 image ‘‘Lena’’ used in the experiments, a watermarked version, and the perceptual mask, obtained from a visibility function defined in [6]. In all the cases considered the empirical values have been obtained watermarking the Lena image with different keys and then averaging out the results. In figure 2 we plot the bit error rate (BER) as a function of the number of pixels per information bit when Wiener filtering is performed before detection to eliminate part of the noise due to the original image and no attack is performed. Figure 3 shows the BER when

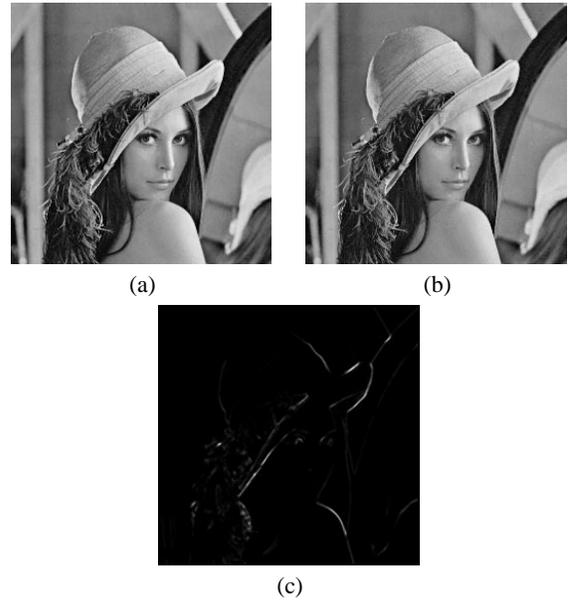


Figure 1: (a) Original image, (b) watermarked version and (c) perceptual mask.

the image is attacked with additive Gaussian noise with variance at each pixel shaped by the perceptual mask to avoid visibility. Figure 4 shows a similar plot for a Wiener filtering attack. Three cases are considered in each plot: uncoded (i.e. binary antipodal), and codes BCH(63,36) and BCH(63,10), whose minimum distances are 11 and 27 respectively. The empirical curves have been obtained by averaging over 50 keys. We can observe that the best code is not the one with the greatest minimum distance, but the one with the least redundancy. We can also observe that in all the cases coding achieves better performance for a number of pixels per information bit greater than a certain minimum amount. The difference between the empirical BER and the analytical upper bound is due to small errors in the estimation of  $a$  and  $\gamma$ .

In figure 5 we show the Chernoff bound for the probability of false alarm ( $P_F$ ) and the probability of detection ( $P_D$ ) without channel coding (i.e. binary antipodal) and with a Golay(23,12) code for message lengths of 60 and 240 bits. We have also plotted the empirical  $P_D$  for each theoretical value of  $P_F$  (note that  $P_F$  is extremely low to be estimated through simulation). The empirical values have been obtained by averaging over 100 keys. We can observe how the ROC degrades as we increase the message length due to the decrease in SNR of the equivalent channel. We can also see how coding degrades the ROC for a fixed message length, due to the decrease in SNR associated with the redundancy introduced by the code. The degradation is more apparent for large message sizes.

### 6. CONCLUSIONS

In this paper we have studied the application of channel coding schemes in a spatial watermarking system for copyright protection of images. We have obtained detector structures and analytical bounds for the BER and the ROC which can be used to know the achievable performance for a given image. These bounds have been contrasted with simulation results performed with several

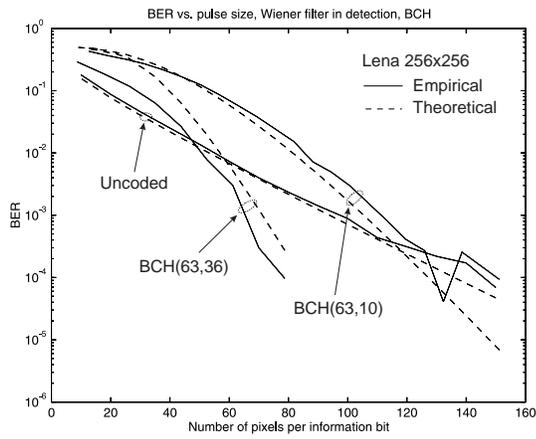


Figure 2: Bit error rate with Wiener filter preprocessing prior to detection.

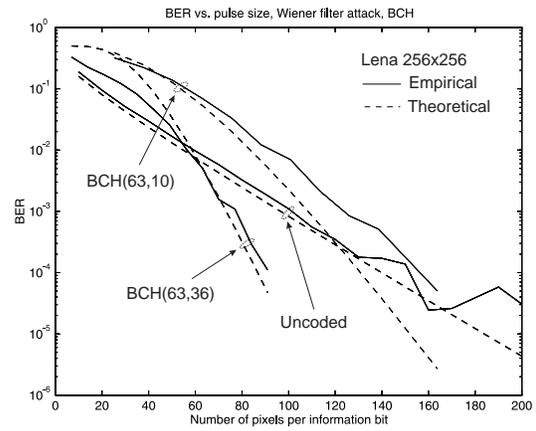


Figure 4: Bit error rate for Wiener filtering attack and Wiener filter preprocessing prior to detection.

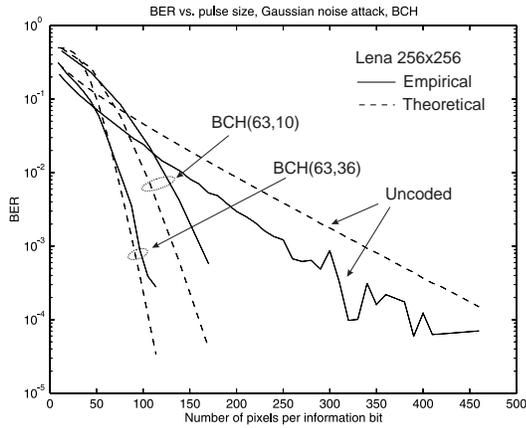


Figure 3: Bit error rate for worst case additive Gaussian noise attack and Wiener filter preprocessing prior to detection.

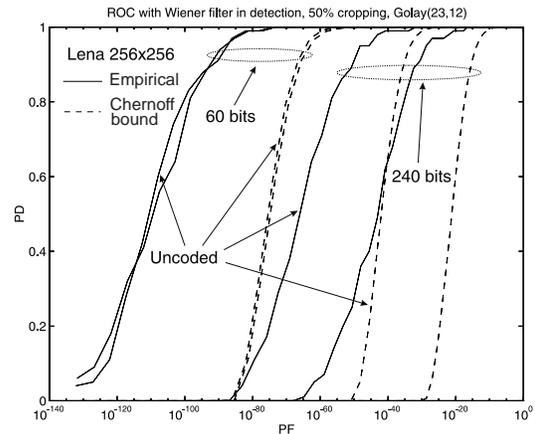


Figure 5: Chernoff bound for the ROC of the watermark detection test with Wiener filtering before detection.

block codes. We observed that the use of block codes results in an improvement of the BER for small bit rates and a degradation for large bit rates. The main factors which determine the difference in performance are the minimum distance and the redundancy of the code. We have also observed that for a given bit rate, coding results in a degraded ROC. These conclusions can be extended to other families of codes.

## 7. REFERENCES

- [1] F. M. Boland, J. J. K. O. Ruanaidh, and C. Dautzenberg, "Watermarking digital images for copyright protection," in *IEE International Conference on Image Processing and its Applications*, (Edinburgh), pp. 326–330, 1995.
- [2] I. J. Cox, J. Kilian, T. Leighton, and T. Shanon, "Secure spread spectrum watermarking for multimedia," Tech. Rep. 95-10, NEC Research Institute, Princeton, NJ, USA, 1995.
- [3] F. Hartung and B. Girod, "Digital watermarking of raw and compressed video," in *Digital Compression Technologies and Systems for Video Communications* (N. Ohta, ed.), vol. 2952, pp. 205–213, SPIE Proceedings Series, October 1996.
- [4] 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," to appear in *IEEE J. Select. Areas Commun.*, april 1998.
- [5] J. R. Hernández, F. Pérez-González, and J. M. Rodríguez, "Data hiding for copyright protection of still images," in *COST 254, Emerging Techniques for Communication Terminals*, (Toulouse, France), ENSEEIHT, July 1997.
- [6] J. S. Lim, *Two-Dimensional Signal and Image Processing*. Prentice-Hall, 1990.
- [7] M. D. Swanson, B. Zhu, and A. H. Tewfik, "Robust data hiding for images," in *Proc. IEEE Digital Signal Processing Workshop*, (Loen, Norway), pp. 37–40, September 1996.
- [8] M. D. Swanson, B. Zhu, and A. H. Tewfik, "Transparent robust image watermarking," in *Proc. IEEE Int. Conf. on Image Processing*, vol. III, (Lausanne, Switzerland), pp. 211–214, September 1996.
- [9] H. L. V. Trees, *Detection, Estimation and Modulation Theory*, vol. Part I. John Wiley & Sons, Inc., 1968.