

# ASYMPTOTICALLY OPTIMAL SCALAR QUANTIZERS FOR QIM WATERMARK DETECTION

Jean-Philippe Boyer\*, Pierre Duhamel\* and Jacques Blanc-Talon†

\* LSS, 3 rue Joliot-Curie, 91190 Gif sur Yvette - France

{jean-philippe.boyer, pierre.duhamel}@lss.supelec.fr

† CTA/GIP, 16 bis av. Prieur de la Côte d'Or, 94114 Arcueil - France

jacques.blanc-talon@etca.fr

## ABSTRACT

This paper investigates asymptotically optimal scalar quantizers to address QIM watermark detection with i.i.d. host data and additive noise. False-alarm probability of detection is chosen as the cost to be minimized, keeping the embedding distortion and the miss probability upper-bounded. To avoid the intractability of false-alarm probability, Kullback distance between watermarked and non-watermarked data is adopted instead. The problem is then to seek the quantizer which maximizes the false-alarm error exponent under distortion constraint. Using Lagrange multiplier minimization, a quantizer updating Lloyd-Max-like procedure is used to solve the optimization. For experimental aspects, host data and noise have been set gaussian. In comparison with uniform or Lloyd-Max quantizers, it turns out that detection performances can be notably enhanced by using proposed application-optimized quantizers. The gain is effective even for small number  $N$  of sample at the detector input. However, this gain becomes more substantial as  $N$  grows. This also emphasises that good quantizers in terms of distortion are not suitable for detection task.

## 1. INTRODUCTION

Quantization Index Modulation (QIM) [1] has been specifically introduced for data-hiding purposes and has demonstrated performance merits. The performances of these schemes are usually evaluated according to achievable transmission rates. Alternatively, it has been shown [2, 3, 4] that QIM systems can also be a relevant choice in zero-bit watermarking (*aka.* one-bit watermarking). This specific problem addresses the situation where one tries to determine whether an arbitrary watermark is embedded into a noisy content. Zero-bit watermarking is of main interest in contexts such as copyright verification [5] or integrity checking [6, 7]. The natural criterion associated to this detection task is the Receiver Operating Characteristic (ROC), *i.e.* the trade-off between the probability to falsely detect the watermark and the probability to miss the presence of the watermark.

In data-hiding context, several works have proposed to design non-uniform quantizers to enhance the robustness of transmission [8, 9]. In this framework, a quantizer coset is to be associated to each transmitted message. For a binary transmission, these studies then design two non-uniform quantizer cosets. Liu [8] has proposed an iterative algorithm to design quantizers which guarantee good trades-off between robustness and distortion. In the joint watermarking and compression context, Wu [9] addresses the prob-

lem to design nearest neighbor quantizers which minimize the binary error probability of transmission.

As opposed to data-hiding applications where the task is to decide between rival transmitted messages, QIM watermark detection issue involves a single quantizer which has to be suitably designed to be able to distinguish watermarked and non-watermarked data. So far, this problem has been only studied for structured quantizers (uniform [3, 4] and nearly spherical Voronoï cell [2]). In these studies, the host signal statistics have not been taken into account due to additional arguments (in instance, projection onto carriers, dithering or high resolution assumption). Nevertheless, the knowledge of the host and noise statistics can naturally provide some performance increasing. Taking into account the latter point, we propose in this paper an iterative method to devise application-specific scalar quantizers which provide ROC performance gain. Our quantizer model does not assume nearest neighbor rule. Performance comparison with uniform, Lloyd-Max (LM) [10] and Distortion Compensated (DC) uniform quantizers and Spread-Spectrum Modulation (SSM) [11] is carried out with i.i.d. gaussian host data and additive white gaussian noise.

## 2. CONTEXT

### 2.1. QIM Watermark Detection Context

Let  $s = \{s_n\}_{1 \leq n \leq N}$  be  $N$  real samples of an i.i.d. host signal distributed according to the component probability density function (pdf)  $p_s(s_n)$ . We denote the host signal variance  $\sigma_s^2 = N^{-1} \mathbb{E} \|s\|^2$ . For the QIM process [1] to be applied, we introduce (*c.f.* Fig. 1) a scalar quantizer  $Q(\cdot)$  defined by a codebook of  $\mathcal{M}$  reproduction levels  $c = \{c_i\}_{1 \leq i \leq \mathcal{M}}$  such as  $c_1 \leq c_2 \leq \dots \leq c_{\mathcal{M}}$ . The  $i^{\text{th}}$  quantization cell  $V_i$  is an interval  $[b_i, b_{i+1})$ , where the non-decreasing set  $b = \{b_i\}_{1 \leq i \leq \mathcal{M}+1}$  is the end points of the quantizer. By construction,  $c_i$  is constrained to belong to  $V_i \cup \{b_{i+1}\}$ ,  $\forall i$ . We require that  $b_1 = -\infty$  and  $b_{\mathcal{M}+1} = \infty$  so that  $\bigcup_{i=1}^{\mathcal{M}} V_i = \mathbb{R}$ . For a quantizer input  $u \in \mathbb{R}$  belonging to cell  $V_i$ , the mapping is classically done by  $Q(u) = c_i$ . The QIM watermark  $w \in \mathbb{R}^N$  is produced component-wise by computing the quantization error  $w_n = Q(s_n) - s_n$ ,  $1 \leq n \leq N$  and the watermark content is defined as  $x = s + w$ . Hence, when  $s_n$  belongs to  $V_i$ , we have  $x_n = c_i$ . The mean embedding distortion defined as  $D_w = N^{-1} \mathbb{E} \|w\|^2$  equals  $\int_{\mathbb{R}} p_s(u) (Q(u) - u)^2 du = \sum_{i=1}^{\mathcal{M}} \int_{V_i} p_s(u) (c_i - u)^2 du$ . Signal  $x$  undergoes an additive i.i.d. noise  $v \in \mathbb{R}^N$  with a component pdf  $p_v(v_n)$ , producing a signal  $r = x + v$ . The noise variance

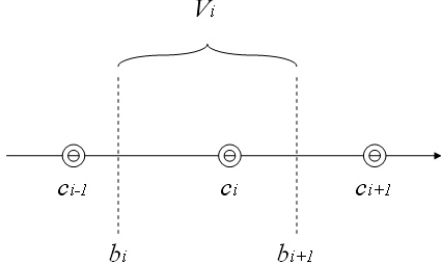


Figure 1: Considered  $\mathcal{M}$ -cell non-uniform quantizer.

is denoted  $\sigma_v^2 = N^{-1} \mathbb{E} \|\mathbf{v}\|^2$ . We then define the Watermark-to-Noise power Ratio by  $wnr = D_w/\sigma_v^2$  and the host Signal-to-Noise power Ratio  $snr = \sigma_s^2/\sigma_v^2$ .

For a signal  $\mathbf{r}$  to be tested, the watermark detection issue (also named zero-bit watermarking) is defined introducing the two following hypotheses:

$$\begin{aligned} H_0 : \quad & \mathbf{r} \text{ is actually not watermarked, } \textit{i.e.} \quad \mathbf{r}|_{H_0} = \mathbf{s} + \mathbf{v}, \\ H_1 : \quad & \mathbf{r} \text{ contains the watermark, } \textit{i.e.} \quad \mathbf{r}|_{H_1} = \mathbf{x} + \mathbf{v}, \end{aligned}$$

and a detector has to decide between both hypotheses without access to the host signal (blind watermarking). Knowing distributions  $p_s$  and  $p_v$  and the quantizer, the optimal detector is a Neyman-Pearson test [12]. This is based on the log-likelihood ratio

$$\begin{aligned} \Lambda(\mathbf{r}) &= \log \frac{Pr(\mathbf{r}|H_1)}{Pr(\mathbf{r}|H_0)} \quad (1) \\ &= \sum_{n=1}^N \log \frac{Pr(r_n|H_1)}{Pr(r_n|H_0)} \triangleq \sum_{n=1}^N \log \frac{p_1(r_n)}{p_0(r_n)}. \quad (2) \end{aligned}$$

(2) holds since all signals are assumed to be i.i.d. For a given detection threshold  $\tau$ , the associated decision rule is

$$\Lambda(\mathbf{r}) \begin{cases} \geq \tau \Rightarrow \text{choose } H_1, \\ < \tau \Rightarrow \text{choose } H_0. \end{cases} \quad (3)$$

The efficiency of this test is measured by the two types of error probabilities: the False-alarm probability to falsely detect the watermark  $P_F(\tau) = Pr(\Lambda(\mathbf{r}) \geq \tau|H_0)$  and the Miss detection probability  $P_M(\tau) = Pr(\Lambda(\mathbf{r}) < \tau|H_1)$ .

## 2.2. Intuitive Aspects

We now give expressions of distributions  $p_0(r_n)$  and  $p_1(r_n)$ , for  $r_n \in \mathbb{R}$ . Under hypothesis  $H_0$ ,  $\mathbf{r} = \mathbf{s} + \mathbf{v}$ .  $\mathbf{s}$  and  $\mathbf{v}$  being statistically independent,  $p_0$  equals the pdf convolution  $p_0(r_n) = \int_{\mathbb{R}} p_v(u) p_s(r_n - u) du$  which is of course independent of the quantizer. Under  $H_1$ ,  $\mathbf{r} = \mathbf{x} + \mathbf{v}$  with  $\mathbf{x}$  and  $\mathbf{v}$  independent. Note however that  $\mathbf{x}$  is a discrete random variable. Hence  $p_1$  has the discrete convolution form  $p_1(r_n) = \sum_{i=1}^{\mathcal{M}} Pr(x_n = c_i) p_v(r_n - c_i)$ . Since the event  $x_n = c_i$  is equivalent so that  $s_n$  lies in  $V_i$ , we then have  $p_1(r_n) = \sum_{i=1}^{\mathcal{M}} Pr(s_n \in V_i) p_v(r_n - c_i)$ . Some illustrations of typical pdf's are depicted on Fig. 2. For watermarked data, it can be seen that the presence of the quantizer reproduction levels scars the shape of the non-watermarked pdf, which provides

a specific statistical signature. Intuitively, the stronger these alterations are, the easier the discrimination between  $p_0$  and  $p_1$  is, but the greater the embedding distortion is.

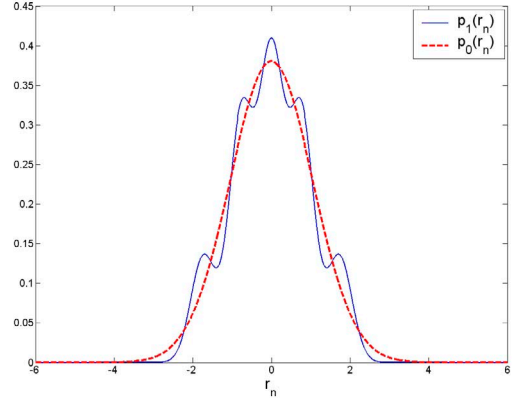


Figure 2: Illustration of the data pdf's under  $H_0$  and  $H_1$  with the 5-cell LM quantizer. Here, host signal and noise have been set gaussian, with  $\sigma_s^2 = 1$ ,  $snr = 10$  dB and a  $wnr = -2.2$  dB.

## 3. CONSIDERED PROBLEM

### 3.1. Problem Statement

As it has been discussed in [7], detection problems in watermarking can be roughly divided in three categories: Copyright systems aim at minimizing the miss-probability [5], keeping the false-alarm rate under an applicative acceptable bound. Integrity checking issues [6] target a minimal false-alarm probability, keeping the miss probability rate under an acceptable level  $P_0$ . Then, a "mix" approach jointly minimizes half the sum of both error probabilities [11]. In this paper, we choose the integrity checking approach. The watermarks then acts as an integrity stamp [6, 7]. Additionally, all these multimedia watermarking scenarii are subjected to a maximal admissible embedding distortion  $D_0$  to avoid noticeable degradations. Thus, for a given number of cells  $\mathcal{M}$ , our problem is to seek a quantizer which fulfills

$$\min_{e,b} P_F \text{ subject to } \begin{cases} P_M \leq P_0 \\ D_w \leq D_0. \end{cases} \quad (4)$$

As opposed to classical source coding contexts, the number of cell  $\mathcal{M}$  is here a free parameter. Note however that  $\mathcal{M}$  has to be large enough to meet the distortion constraint. Indeed, denote the minimal distortion  $D_{min}(\mathcal{M})$  induced by a  $\mathcal{M}$ -cell scalar quantizer in quantizing source  $s$ . For a given  $D_0$ ,  $\mathcal{M}$  must be chosen such as

$$D_{min}(\mathcal{M}) \leq D_0 \quad (5)$$

otherwise we have  $D_0 < D_{min}(\mathcal{M}) \leq D_w$  and there cannot be any quantizer which verifies the distortion constraint. Since  $D_{min}(\mathcal{M})$  is clearly decreasing with respect to  $\mathcal{M}$ , tending to 0<sup>1</sup>, one can always find  $\mathcal{M}$  sufficiently large which fulfills (5). For practical cases, if the logarithm of  $p_s$  is concave (which is the case for laplacian and gaussian laws), it is known [10] that the LM algorithm yields a globally optimal quantizer. In this case,  $D_{min}(\mathcal{M})$  is thus the distortion induced by the LM quantizer.

<sup>1</sup>For example, the high resolution (*i.e.*  $\mathcal{M} \gg 1$ ) distortion equals  $D_{min}(\mathcal{M}) = \frac{\pi\sqrt{3}}{2} \sigma_s^2 \mathcal{M}^{-2}$  in choosing  $s$  gaussian [10].

### 3.2. Proposed Cost Criterion

False-alarm probability is an intractable function of  $\mathbf{b}$  and  $\mathbf{c}$ . Instead, we resort to the Kullback-Leibler distance [13] as an alternative cost criterion. This non-negative dissimilarity measure between distributions is defined by

$$D(p_1||p_0) = \int_{\mathbb{R}} p_1(u) \log \frac{p_1(u)}{p_0(u)} du. \quad (6)$$

This quantity is known to capture the difficulty to discriminate between hypotheses  $H_0$  and  $H_1$  [14]. The larger is  $D(p_1||p_0)$ , the smaller  $P_F$  is expected to be. More precisely, the relevance of this criterion can be formalized by the Stein's lemma [13] which states that, for a fixed upper-bound  $P_0$  on  $P_M$  uniformly over all  $N$ ,  $P_F$  decays exponentially in the number of observations  $N$  at rate  $D(p_1||p_0)$ . Defining the false-alarm error exponent  $E_F = \lim_{N \rightarrow \infty} -\frac{1}{N} \log P_F$ , this means that  $E_F = D(p_1||p_0)$ . Hence, we now seek the quantizer which maximizes this false-alarm decay rate. This criterion then provides a form of asymptotical optimality. However, it will be demonstrated that the gain provided by this approach is effective even for small  $N$ . Our problem reduces to

$$\max_{\mathbf{c}, \mathbf{b}} D(p_1||p_0) \quad \text{subject to } D_w \leq D_0. \quad (7)$$

### 4. OPTIMIZATION

We propose to solve (7) using a Lagrange multiplier approach. Let be the Lagrangian function  $\mathcal{L}$  defined by  $\mathcal{L}(\lambda) = -D(p_1||p_0) + \lambda D_w$  for some positive  $\lambda$ . For a given  $\lambda$  and for a given number of cells  $\mathcal{M}$  satisfying condition (5), we minimize  $\mathcal{L}(\lambda)$  by an iterative descent LM-like algorithm. Our method is similar to the one proposed in [9] in joint watermarking and compression data-hiding context. Nevertheless, contrary to this study, we do not apply the nearest neighbor rule. The procedure is divided in two steps: first, we begin by fixing the end points  $\mathbf{b}$  and seek the reproduction levels  $\mathbf{c}$  which minimize  $\mathcal{L}(\lambda)$ . To that end, we proceed independently level by level: we fix all levels except one, say the  $i^{\text{th}}$ . Then we seek the best position of  $c_i$  in  $V_i$  which minimizes  $\mathcal{L}(\lambda)$ . In practice, this optimization task is solved numerically. Once this is done, we do the same procedure for the  $i^{\text{th}} + 1$  level and so on until all levels have been updated. At the second step, we fix  $\mathbf{c}$  and minimize  $\mathcal{L}(\lambda)$  over  $\mathbf{b}$  following the same updating strategy. These two steps are alternatively iterated until the criterion variation drops below a prescribed threshold  $\epsilon$ . The criterion is then decreasing. If  $\mathcal{L}$  is lower-bounded, the convergence is guaranteed. The term  $\lambda D_w$  of  $\mathcal{L}$  is clearly positive so it is sufficient to find an upper bound on  $D(p_1||p_0)$ . For practical implementation of this procedure, one should set the extremal end points  $b_1 = -U$  and  $b_{\mathcal{M}+1} = U$  to compute integrals terms, where  $U$  is an arbitrarily large positive real number with respect to the width of distributions  $p_0$  and  $p_1$  so that ensuring that most of area of these pdf's are contained in  $[-U, U]$ .  $p_v$  being defined on the closed interval  $[-U, U]$ , it exists some  $A_1 > 0$  such that  $p_v(u) \leq A_1$  for  $u \in [-U, U]$ . Thus,  $p_1$  can be upper-bounded by  $p_1(u) = \sum_{i=1}^{\mathcal{M}} Pr(s_n \in V_i) p_v(u - c_i) \leq A_1 \sum_{i=1}^{\mathcal{M}} Pr(s \in V_i) = A_1$ . Last equality holds since cells  $\{V_i\}$  map the real line. Assume additionally that the support of  $p_0$  is empty over  $[-U, U]$  so that it exists  $A_2 > 0$  such that  $p_0(u) \geq A_2$  for  $u \in [-U, U]$ . Using these two latter boundings on  $p_0$  and  $p_1$ , we straightly have  $D(p_1||p_0) = \int_{-U}^U p_1(u) \log \frac{p_1(u)}{p_0(u)} du \leq 2UA_1 \log \frac{A_1}{A_2}$ . Hence, the procedure converges in practice. By sweeping  $\lambda$  over the range

from zero to infinity, we finally find the appropriate multiplier  $\lambda_0$  which meets the distortion constraints  $D_w = D_0$ . Note that for large  $\lambda$ ,  $\mathcal{L}$  is dominated by the distortion constraint and the result of the optimization is then the LM quantizer.

### 5. RESULTS

For experimental aspects, we choose  $p_s$  and  $p_v$  gaussian with  $p_s \sim \mathcal{N}(0, 1)$ . Note however that this procedure can be straightly implemented with a large variety of statistics. In the sequel, we set  $unr = 0$  dB and maximal distortion  $D_0 = 0.07995$ . Doing so, we have  $snr = 10.97$  dB. The extremal end points are fixed setting  $U = 6.365$ . These specific values have been chosen so that the comparison with simple quantizers (uniform and LM quantizers) is fair. Indeed, the distortion of the 13-cell uniform quantizer and the 5-cell LM quantizer equal both  $D_0$ . Hence, the minimal value of  $\mathcal{M}$  which fulfills condition (5) is 5. For given  $\mathcal{M} \geq 5$  and  $\lambda \geq 0$ , we initialize the procedure with the  $\mathcal{M}$ -cell uniform quantizer and we stop the minimization of  $\mathcal{L}(\lambda)$  when its variation drops under  $\epsilon = 0.001$ . The appropriate  $\lambda_0$  which meets the distortion constraint is fixed with a relative error lower than 0.0001. Optimization results are depicted on Fig. 3. It can be seen that  $D(p_1||p_0)$  is increasing with respect to  $\mathcal{M}$  for  $\mathcal{M} \leq 9$ . When taking  $\mathcal{M}$  greater than 9, optimized quantizers provide equivalent scores ( $D(p_1||p_0) \simeq 0.147$ ). As expected, it turns out that uniform and LM quantizers provide lower Kullback distances ( $D(p_1||p_0) = 0.049$ , respectively 0.042).

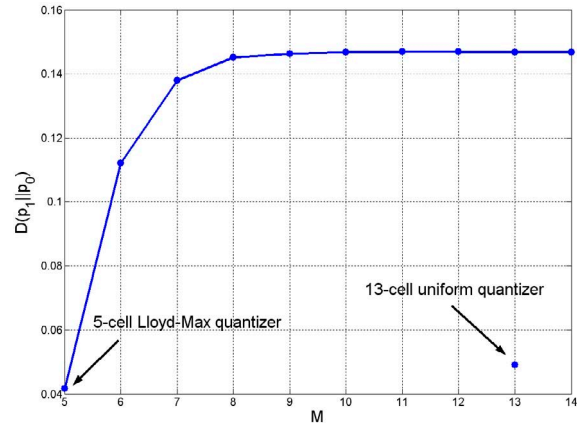


Figure 3: Result of the maximization of  $D(p_1||p_0)$  with respect to the number of quantizer cells  $\mathcal{M}$ , keeping the distortion constant at  $D_0$ . The first point of the curve ( $\mathcal{M} = 5$ ) corresponds to the LM quantizer. The single point in the right down corner depicts the score of uniform quantizer.

We now propose to assess the effective gain in term of ROC performance. For this study, we work with the 10-cell optimized quantizer. In order to reliably estimate the trade-off between false-alarm and miss probabilities, we use large deviation techniques. These techniques are clearly presented in [12], chap. II.7 and used in [4]. Comparisons with uniform, ML and uniform DC quantizers and SSM are provided. As expected, for  $N = 128$ , it can be seen on Fig. 4 that the optimized quantizer version of the scheme notably outperforms rival schemes with no DC's. This underlines that good scalar quantizers in terms of distortion (LM quantizers)

and uniform quantizer are not well adapted for watermark detection. Such as DC, use of optimized quantizers is an effective technique to enhance performances with several orders of magnitude. Of course, these two techniques can be jointly applied.

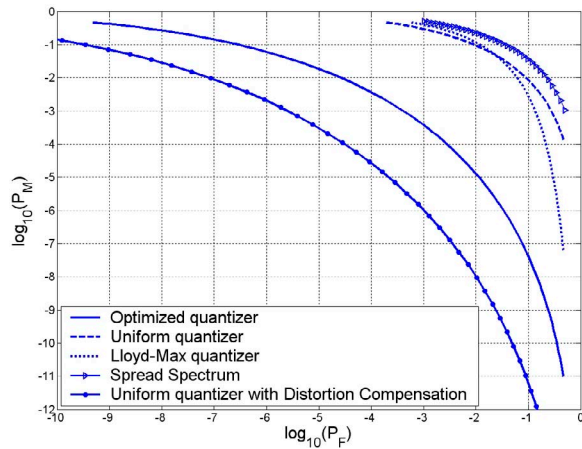


Figure 4: ROC profile comparison when using Optimized, Uniform, LM and DC Uniform quantizers and SSM (with  $N = 128$ ,  $wnr = 0$  dB and  $snr = 10.97$  dB).

The Kullback distance criterion (6) provides a form of asymptotical optimality as  $N$  grows. Thus, performance gain for small  $N$  is not guaranteed. In order to evaluate this gain rate, we propose to compute performance with respect to  $N$ . To illustrate this point, we now work with the overall probability of detection error, assuming that  $H_0$  and  $H_1$  have equal priors and equal costs. This quantity is defined as  $P_E = P_F(\tau = 0)/2 + P_M(\tau = 0)/2$ , which is more compact but captures the global shape of ROC curves. It is shown on Fig. 5 that the gain is effective even for small values of  $N$ . However, the gain is more substantial when taking  $N$  large since the decreasing rate of  $P_E$  for the optimized quantizer is greater than the ones of rival schemes.

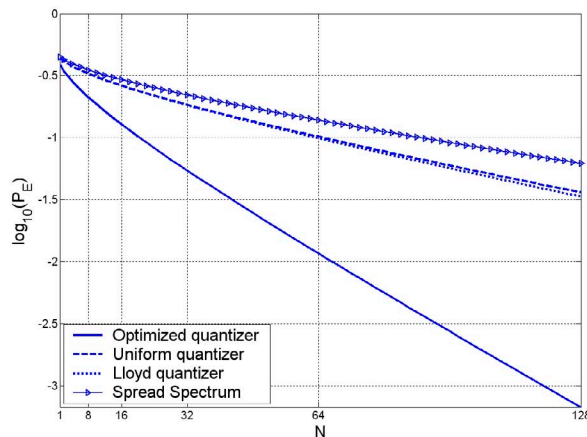


Figure 5: Overall probability of error  $P_E$  with respect to  $N$  when using Optimized, Uniform and LM quantizers and SSM (with  $wnr = 0$  dB and  $snr = 10.97$  dB).

## 6. CONCLUSION AND PERSPECTIVES

We have investigated asymptotically optimal scalar quantizers to address QIM watermark detection with i.i.d. host data and additive noise. False-alarm probability of detection has been chosen as the cost to be minimized, keeping the embedding distortion and the miss probability upper-bounded. To avoid the intractability of false-alarm probability, Kullback distance between watermarked and non-watermarked data has been adopted instead. The problem is then to maximize the false-alarm error exponent under distortion constraint. Using Lagrange multiplier minimization, a quantizer updating Lloyd-Max-like procedure has been used to solve the optimization. Experimentally, host data and noise have been set gaussian. In comparison with uniform or Lloyd-Max quantizers, it has turned out that ROC performances can be notably enhanced by using proposed application-optimized quantizers. This gain increases with respect to the number of sample at the detector input. This emphasises that good quantizers in terms of distortion are not suitable for detection task. Note that distortion compensation principle is readily applicable. This would potentially provide significant performance increasing [4]. The proposed procedure can be straightly applied with different typical host and noise statistics. Laplacian host and uniform noise could be an interesting case.

## 7. REFERENCES

- [1] B. Chen and G. Wornell. Quantization index modulation: a class of provably good methods for digital watermarking and information embedding. *IEEE Trans. Inform. Theory*, 47:1423–1443, May 2001.
- [2] T. Liu and P. Moulin. Error exponents for one-bit watermarking. In *Proc of IEEE ICASSP'03*, volume 3, Hong Kong, April 2003.
- [3] L. Pérez-Freire, P. Comesaña-Alfaro, and F. Pérez-González. Detection in quantization-based watermarking: performance and security issues. In *SPIE, Watermarking of Multimedia Contents VII*, San Jose, California, USA, January 2005.
- [4] J.P. Boyer, P. Duhamel, and J. Blanc-Talon. Performance analysis of scalar dc-qim for watermark detection. In *Proc. of ICASSP'06, to appeared*, Toulouse, France, May 2006.
- [5] J. R. Hernandez and F. Perez-Gonzalez. Statistical analysis of watermarking schemes for copyright protection of images. *Proc. IEEE*, 87:1142–1166, July 1999.
- [6] J. P. Boyer, P. Duhamel, and J. Blanc-Talon. Game-theoretic analysis of a semi-fragile watermarking scheme based on scs. In *Proc. of ICIP'05*, Genova, Italy, September 2005.
- [7] J.P. Boyer, P. Duhamel, and J. Blanc-Talon. Spread-spectrum substitution watermarking game. In *Proc. of ICME'06, to appeared*, Toronto, Canada, July 2006.
- [8] N. Liu and K. P. Subbalakshmi. Non-uniform quantizer design for image data hiding. In *Proc. of ICIP'04*.
- [9] G. Wu and E. H. Yang. Joint watermarking and compression using scalar quantization for maximizing robustness in the presence of additive gaussian attacks. *IEEE Transactions on Signal Processing*, 53(2):834–844, February 2005.
- [10] R. M. Gray and D. L. Neuhoff. Quantization. *IEEE Trans. on Information Theory*, 44(6):2325–2383, October 1998.
- [11] P. Moulin and A. Ivanovic. The zero-rate spread-spectrum watermarking game. *IEEE Trans. on S. P.*, 51(4):1098–1117, April 2003.
- [12] H. L. Van Trees. *Detection, Estimation, and Modulation Theory, part I*. John Wiley & Sons, Inc.
- [13] T.M. Cover and J.A. Thomas. *Elements of Information Theory*. John Wiley and Sons, Inc., 1991.
- [14] C. Cachin. An information-theoretic model for steganography. *Information and Computation*, 192(1):41–56, July 2004.