

PERCEPTUAL WATERMARKING OF NON I.I.D. SIGNALS BASED ON WIDE SPREAD SPECTRUM USING SIDE INFORMATION

Gaëtan Le Guelvouit, Stéphane Pateux and Christine Guillemot

IRISA/INRIA, Campus de Beaulieu,
35042 Rennes Cedex, FRANCE

Tel: +33 2 99 84 73 60; fax: +33 2 99 84 25 31

e-mail: Gaetan.Le_Guelvouit@irisa.fr

ABSTRACT

The theoretical foundations of data hiding have been revealed by formulating the problem as message communication over a noisy channel, i.e. the host document. In light of this definition, some solutions have been proposed. Unfortunately, the performances of those methods are limited due to interference with the host signal. Considering spread spectrum information hiding with non i.i.d. Gaussian host signals and weighted distortion measures, we propose in this paper a game-theoretic resolution of the problem using side information.

1. INTRODUCTION

Digital watermarking consists in embedding an invisible message within a host signal. This paper deals with the problem of a blind, robust and symmetric data hiding scheme. In light of information and communication theory, this problem can be seen as reliably transmitting a message over a noisy channel, noise being due to modification or attacks of the host document. The attacks are often modeled as the addition of white Gaussian noise (AWGN channel) [1, 2], or as linear filtering plus additive noise [3, 4]. The perceptual sensitivity of the host signal is often taken into account for choosing embedding sites and strength.

After presenting the watermarking general scheme in Sec. 2, we introduce side information principles in Sec. 3. Then digital watermarking is revisited in light of a general channel model. While the host signal is often considered as an ergodic wide sense stationary Gaussian random process, which is rarely satisfied for real signals, we assume that it can be modeled as a set of independent non identically distributed Gaussian random variables (non i.i.d. signals). The attack channel is considered to be a scaling¹ and an addition of white Gaussian noise (SAWGN). The game-theoretic resolution of the problem (Sec. 4) leads to closed-form expressions of the optimal embedding and extraction parameters, leading to a practical watermarking scheme.

¹In the case of embedding in the Fourier domain, convolutional filtering can be seen as a kind of scaling.

2. PROBLEM STATEMENT

Many approaches introduced so far assume that the signal \underline{x} can be modelled as an ergodic zero-mean wide sense stationary Gaussian random process [2, 3]. This assumption is rarely satisfied for real signals or for content adaptive watermarks. We assume instead that the host signal \underline{x} can be modelled as the realization of a set of non stationary Gaussian random variables $\underline{X} = \{X_1, X_2, \dots, X_m\}$ where $X_i \sim \mathcal{N}(0, \sigma_{X_i}^2)$. Let $\underline{b} = \{b_1, b_2, \dots, b_n\}$ be the message to be embedded in the host signal. The information is then used as a key for indexing pseudo-random noise sequences which are additively combined with the signal, weighted by α_i factors. Let \underline{G} be a $n \times m$ matrix composed of n pseudo-random vectors $\underline{G}_j \in \{-1, +1\}^m$. To get less distortion without loss of performance, the signal is filtered after embedding with a Wiener filter (i.e. a scaling factor defined by $\gamma_i^w = \frac{\sigma_{X_i}^2}{\sigma_{X_i}^2 + n\alpha_i^2}$, since $\sigma_{W_i}^2 = n\alpha_i^2$). The watermarked signal is obtained by

$$y_i = \gamma_i^w [x_i + w_i] = \gamma_i^w \left[x_i + \alpha_i \sum_{j=1}^n G_{i,j} b_j \right], \quad (1)$$

where x_i represents the i^{th} site of the host signal and y_i the corresponding watermarked site. In order to extract each embedded bit b_j , a correlation product between the vector \underline{G}_j and \underline{y} is generally computed. The term α_i is a weighting factor allowing to adjust the amplitude (or energy) of the mark for robustness and low perceptual visibility.

The attack channel is often assumed to be AWGN [1, 2]. This model assumes that the distortion induced by the attack is independent of the watermarked signal, hence can hardly apply to attacks such as filtering and compression. More accurate models assuming that the distortion depends on the watermarked signal and based on linear filtering plus additive noise have been considered in [3, 4]. Here, we consider

that the attacked signal \underline{y}' can be expressed as

$$y'_i = \gamma'_i y_i + \delta_i = \gamma_i x_i + \gamma_i \alpha_i \sum_{j=1}^n G_{i,j} b_j + \delta_i, \quad (2)$$

where $\gamma'_i = \gamma_i / \gamma_i^w$ is an attenuation factor on each watermarked site. This amounts to consider the attack channel as an SAWGN channel (amplitude scaling by the factor γ'_i and additive white Gaussian noise of $\delta_i \sim \mathcal{N}(0, \sigma_{\delta_i}^2)$).

The distortion measure is defined as a weighted sum of the MSE on each sample of the host signal, in order to reflect the perceptual quality. The expected embedding distortion is therefore given by

$$D_{xy} = \sum_{i=1}^m \varphi_i^2 \frac{\sigma_{X_i}^2 \sigma_{W_i}^2}{\sigma_{X_i}^2 + \sigma_{W_i}^2}, \quad (3)$$

where φ_i is a perceptual factor (e.g. see [5]). Similarly, the expected attack distortion is

$$D_{xy'} = \sum_{i=1}^m \varphi_i^2 \left(\sigma_{X_i}^2 (1 - \gamma_i)^2 + \gamma_i^2 \sigma_{W_i}^2 + \sigma_{\delta_i}^2 \right). \quad (4)$$

3. WATERMARKING WITH SIDE INFORMATION

Let us consider an i.i.d. host signal $X \sim \mathcal{N}(0, Q)$, the watermark W with a bounded power constraint $\frac{1}{m} \sum_{i=1}^m W_i^2 \leq P$, and additive noise $Z \sim \mathcal{N}(0, N)$. Costa [6] has shown that the capacity limit in this case of AWGN channel (i.e. $Y = X + W + Z$) is given by

$$C = \frac{1}{2} \log_2 \left[1 + \frac{P}{N} \right]. \quad (5)$$

This limit can be reached with the introduction of additional information $U \sim \mathcal{N}(0, P + \alpha^2 Q)$ such as $U = W + \alpha X$, known both from the encoder and the extractor. The capacity can then be written as

$$C = \max_{\alpha} \{R(\alpha)\} = \max_{\alpha} \{I(U; Y) - I(U; S)\}. \quad (6)$$

The maximum of the previous functional is obtained for $\alpha = \frac{P}{P+N}$, which leads to Eqn. (5). This method is known as watermarking with side information (i.e. SI).

In order to be able to watermark the host signal efficiently (i.e. guaranty that the correct message can be extracted), it is necessary to have $\alpha \geq 1/2$, i.e. $P \geq N$ (see detailed explanations in [7]). This condition may not be satisfied in practical situations. This property can however be verified by embedding in a linear subspace of the original space representation. When considering a subspace with $\epsilon \times m$ components, the allowed embedding distortion on this subspace is P/ϵ , where ϵ represents the ratio between the dimensions of the original space m and the selected subspace.

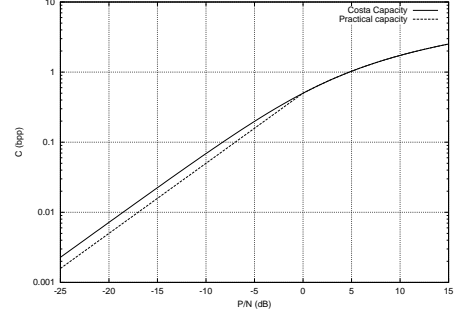


Fig. 1. Comparison between Costa's capacity (Eqn. (5)) and the achievable capacity (Eqn. (8)).

If this subspace is not known to the attacker, the added noise Z will be spread over all the components. Thus the noise on the subspace will still be of the form $\mathcal{N}(0, N)$. Using a linear subspace then permits to artificially increase the SNR P/N by a factor of $1/\epsilon$. The capacity per sample of the original space is then given by

$$C(\epsilon) = \frac{\epsilon}{2} \log_2 \left[1 + \frac{P}{\epsilon N} \right]. \quad (7)$$

Under the constraint that $P/\epsilon \geq N$ (i.e. $\alpha \geq 1/2$), this capacity is maximal for $\epsilon = P/N$. It is then expressed as

$$C = \begin{cases} \frac{1}{2} \log_2 \left[1 + \frac{P}{N} \right] & \text{if } P > N, \\ \frac{1}{2} \log_2 \left[1 + \frac{P}{2N} \right] & \text{otherwise.} \end{cases} \quad (8)$$

Fig. 1 shows the difference between the capacity defined by Costa and the capacity defined by Eqn. (8) that uses a subspace. The loss in term of performance is very low, and the practical capacity is nearly proportional to the capacity defined by Costa: for low SNR, $C_{\text{Costa}} \simeq \frac{1}{2 \ln 2} \frac{P}{N}$. The practical capacity is then reduced by a factor of $\ln 2$, i.e. by 30%.

4. WATERMARKING BASED ON WSS OF NON I.I.D. SIGNALS

As shown in the previous section, using a linear subspace permits to embed reliably a message in any cover signal while reaching performance of side information. Since WSS techniques use a set of pseudo-random vectors to embed the watermark, they can be interpreted as a special case of linear subspace embedding technique, the linear subspace being the one spanned by the weighted vectors \underline{G}_j . When writing the equation

$$y_i = x_i + \alpha \sum_{j=1}^n b_j G_{i,j}, \quad (9)$$

the terms b_j are then to be considered as the components of the watermark embedded in the linear subspace. The terms

b_j are not bits defining the message independently from the host data but, in the spirit of side information, are defined as $\underline{b} = \underline{U} - \alpha_{\text{Costa}} \underline{X}$. ST-DM [8] and ST-SCS [9] are examples of such linear subspace embedding with side information.

4.1. Optimal linear extractor

Now in order to extend those results to non i.i.d. signals, we look for embedding techniques using WSS with side information. Since host data is not ergodic and that perceptual distortion is considered, non white spread transforms will be considered (i.e. α non constant for each site). Developments made in [10] are then extended to the case of side information. Considering embedding Eqn. (1) and SAWGN attacks, the received signal can be expressed as $y'_i = \gamma'_i y_i + \delta_i$, where $\delta_i \sim \mathcal{N}(0, \sigma_{\delta_i}^2)$. Let us consider the global extractor as a composition of a linear transform followed by a watermark extractor in the new linear subspace representation. Let the linear transform be expressed as

$$\hat{b}_j = \sum_{i=1}^m \beta_{i,j} y'_i. \quad (10)$$

In the case of classical WSS technique, the performance is defined as the ratio between the energy of the embedded mark and the sum of all the interference energies [10]. When considering watermarking with side information, the host signal does not influence the performance. The SNR is then defined as the ratio between the energy of the mark and the added noise. This ratio in the subspace allows to estimate the embedding capacity (see Eqn. (8)). The signal to noise ratio E_b/N_0 is therefore expressed as

$$\frac{E_b}{N_0} = \frac{(\sum_{i=1}^m \gamma_i \alpha_i \beta_{i,j} G_{i,j})^2}{\sum_{i=1}^m \beta_{i,j}^2 [a \gamma_i^2 \sigma_{X_i}^2 + b \gamma_i^2 (n-1) \alpha_i^2 + \sigma_{\delta_i}^2]}, \quad (11)$$

where the terms $a \in [0; 1]$ and $b \in [0; 1]$ have been introduced in order to take into account or not respectively the host data interference, and the self-interference of the watermark. Since $E[b_j^2] = 1$, the energy of the embedding process is then expressed by the terms α_i .

We then search for parameters $\beta_{i,j}$ that maximize the extractor performance E_b/N_0 . The derivative of Eqn. (11) with respect to $\beta_{i,j}$ are then computed. This leads² to the optimal value

$$\beta_{i,j} \propto \frac{\alpha_i \gamma_i G_{i,j}}{a \gamma_i^2 \sigma_{X_i}^2 + b \gamma_i^2 (n-1) \alpha_i^2 + \sigma_{\delta_i}^2}, \quad (12)$$

and the corresponding signal to noise ratio

$$\frac{E_b}{N_0} = \sum_{i=1}^m \frac{\alpha_i^2 \gamma_i^2}{a \gamma_i^2 \sigma_{X_i}^2 + b \gamma_i^2 (n-1) \alpha_i^2 + \sigma_{\delta_i}^2}. \quad (13)$$

²See [7] for details.

4.2. Game-theoretic resolution

Given a maximal allowed distortion, the attacker wants to minimize the extractor performance E_b/N_0 . This corresponds to search the parameters γ'_i (i.e. γ_i) and σ_{δ_i} that minimize the functional

$$J_\lambda = \frac{E'_b}{N_0} + \lambda [D_{xy'} - D_{xy'}^{\max}] \quad \text{with} \quad \frac{E'_b}{N_0} = n \frac{E_b}{N_0}. \quad (14)$$

The study of this cost function leads to three domains (see Fig. 2). First, \mathcal{D}_E is defined as the domain where $\sigma_{W_i} > \varphi_i \sqrt{\lambda} \sigma_{X_i}^2$. The optimal attack parameters for this case are $\gamma_i = 0$ and $\sigma_{\delta_i}^2 = 0$ (the data are erased). The second one is \mathcal{D}_W , defined by $\sigma_{W_i} < \varphi_i \sqrt{\lambda} \gamma_i^w (a \sigma_{X_i}^2 + b(n-1) \alpha_i^2)$. The parameters are then $\gamma_i = \gamma_i^w$ and $\sigma_{\delta_i}^2 = 0$ (the data are Wiener filtered). The last domain \mathcal{D}_I corresponds to the optimal attack parameters

$$\begin{cases} \gamma_i &= \frac{\sigma_{X_i}^2 - \frac{\sigma_{W_i}}{\varphi_i \sqrt{\lambda}}}{(1-a) \sigma_{X_i}^2 + \sigma_{W_i}^2 (1-b \frac{n-1}{n})} \\ \sigma_{\delta_i}^2 &= \gamma_i (\gamma_i^w - \gamma_i) (\sigma_{X_i}^2 + \sigma_{W_i}^2). \end{cases} \quad (15)$$

Let us consider the case where $a = b = 0$ (real side information). The Wiener attack is not defined for this case. Given the above attack, the defender strategy is to maximize E_b/N_0 under a constraint of a maximal distortion, i.e. to maximize the cost function $J_\chi = J_\lambda - \chi [D_{xy} - D_{xy}^{\max}]$. For the domain \mathcal{D}_I , this corresponds to the functional for each site

$$J_{\chi|i} = \varphi_i^2 (\lambda - \chi) \gamma_i^w \sigma_{W_i}^2 + 2 \varphi_i \sqrt{\lambda} \gamma_i^w \sigma_{W_i} + \gamma_i^w - 1. \quad (16)$$

Study of the derivative of the functional with respect to σ_{W_i} leads to the optimal embedding energy, given by

$$\sigma_{W_i} = \frac{\sqrt{(\varphi_i^2 (\lambda - \chi) \sigma_{X_i}^2 - 1)^2 + 4 \varphi_i^2 \lambda \sigma_{X_i}^2}}{2 \varphi_i \sqrt{\lambda}} + \frac{\varphi_i^2 (\lambda - \chi) \sigma_{X_i}^2 - 1}{2 \varphi_i \sqrt{\lambda}} \quad (17)$$

Since the other domain \mathcal{D}_E corresponds to the erasure of the data (then the optimal embedding strength factor is 0), parameters from Eqn. (17) are always optimal. We can also deduce a new extractor, defined by $\beta_{i,j} = \varphi_i G_{i,j}$, that permits optimal extraction without any knowledge of attack parameters. Fig. 3 shows the shape of the strength σ_W of the watermark in terms of the host data standard deviation σ_X , for a MSE distortion. We can remark that all the sites are watermarked, and the higher σ_{X_i} is, the stronger it is marked.

5. RESULTS

Performance of the described scheme can be quantified by E'_b/N_0 . The capacity per pixel is defined as $\frac{E'_b}{2N_0}$. In the

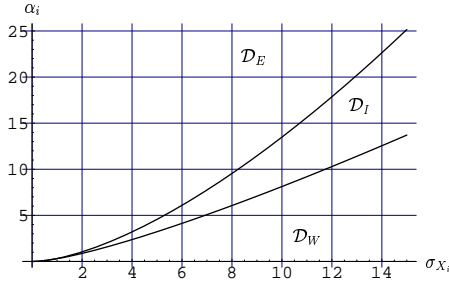


Fig. 2. The tree domains, corresponding to the study of the attack cost function ($a = b = 1$).

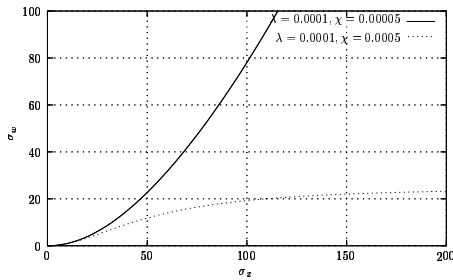
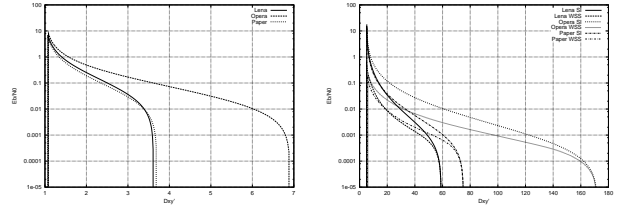


Fig. 3. Insertion strength σ_w in terms of σ_x for optimum watermarking defined by Eqn. (17), with MSE distortion (i.e. $\varphi_i = 1$).

following results, host signal \underline{x} is obtained with a wavelet transform of a gray levels image. Fig. 4 depicts those performances for three common 512×512 images: *Lena*, *Opera* and *Paper*. The performance decreases with the strength of the attack and leads to a capacity of 0 bit when all the watermarked wavelet coefficients are set to zero by the attack. Fig. 5(b) shows the comparison between WSS watermarking with and without side information. As expected, the scheme using side information provides higher SNR than the one without SI. Then higher capacities or lower bit errors rate can be obtained.

6. CONCLUSION

We provided in this paper a solution for watermarking of non i.i.d. signals with side information. Considering weighted distortion in order to better take account of human perception, we analyzed a practical scheme based on spread spectrum. This led to closed-form expressions of embedding and extraction parameters, revealing a practical information hiding scheme. Provided results showed the performance of the whole system and the gain compared to classical WSS technique.



(a) Theoretical performance with psycho-visual factor $\varphi_i = (1 + \sigma_{X_i})^{-\frac{1}{2}}$. Average embedding distortion D_{xy} is set to 1.

(b) Comparison between WSS with and without side information. The average embedding distortion is set to 5 (without psycho-visual factor, i.e. $\varphi_i = 1$).

Fig. 4. Theoretical performances of watermarking.

7. REFERENCES

- [1] S. Servetto, C. I. Podilchuk, and K. Ramchandran, "Capacity issues in digital image watermarking," in *Proc. Int. Conf. on Image Processing*, Chicago, IL, Oct. 1998, vol. 1, pp. 445–449.
- [2] P. Moulin and J. A. O'Sullivan, "Information-theoretic analysis of watermarking," in *Proc. Int. Conf. on Acoustic, Speech and Signal Processing*, Istanbul, Turkey, Jun. 2000.
- [3] J. K. Su, J. J. Eggers, and B. Girod, "Analysis of digital watermarks subjected to optimum linear filtering and additive noise," *IEEE Trans. Signal Proc.: Special Issue on Information Theoretic Issues in Digital Watermarking*, vol. 81, no. 6, Jun. 2001.
- [4] P. Moulin and A. Ivanovic, "The watermark selection game," in *Proc. Conf. on Info. Sciences and Systems*, Mar. 2001.
- [5] A. B. Watson, "DCT quantization matrices visually optimized for individual images," *Proc. SPIE*, vol. 1913, pp. 202–216, 1993.
- [6] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. Info. Thy*, vol. 29, no. 3, pp. 439–441, May 1983.
- [7] S. Pateux, G. Le Guelvouit, and C. Guillemot, "Information-theoretic analysis of WSS watermarking of non i.i.d. Gaussian signals," *submitted to IEEE Trans. Signal Proc.*, Dec. 2001.
- [8] B. Chen and G. W. Wornell, "Quantization index modulation: a class of provably good methods for digital watermarking and information embedding," *IEEE Trans. Info. Thy*, vol. 47, no. 4, pp. 1423–1443, May 2001.
- [9] J. K. Su, J. J. Eggers, and B. Girod, "Performance of a practical blind watermarking scheme," in *Electronic Imaging 2001*, SPIE., Ed., San Jose, CA, Jan. 2001.
- [10] G. Le Guelvouit, S. Pateux, and C. Guillemot, "Information-theoretic resolution of perceptual WSS watermarking of non i.i.d. Gaussian signals," *submitted to European Signal Proc. Conference*, Sep. 2002.