Construction de codes pour tatouage avec prise en compte de l'information adjacente

Gaëtan LE GUELVOUIT, Stéphane PATEUX, Jonathan DELHUMEAU

IRISA/INRIA, Campus de Beaulieu
35042 Rennes Cedex, France
{gleguelv, spateux, jdelhume}@irisa.fr

Résumé – Les articles les plus récents concernant le tatouage l'assimilent à un problème de transmission sur un canal avec information adjacente disponible lors de l'insertion (une partie du bruit est parfaitement connue au moment de l'encodage). Les travaux de Costa ont exposé un schéma théorique, basé sur la construction d'un dictionnaire structuré, permettant de prendre en compte cette information et d'obtenir de bien meilleures performances. Nous présentons dans cet article une technique de construction s'appuyant sur des codes correcteurs poinçonnés. Les résultats obtenus sur le tatouage d'images montrent un apport significatif vis-à-vis du tatouage additif classique.

Abstract – Latest articles consider watermarking as a problem of communication over noisy channel with side information (*i.e.* part of the noise is pefectly known at the encoder). In his work, Costa exposed a theoretical scheme, based on a structured codebook, that takes into account this side information and allows better performances. In this paper, we introduce the construction of such a codebook, based on punctured error correcting codes. The results of side informed image watermarking using our codebook construction demonstrate it's interest over standard additive watermarking techniques.

Introduction

Le développement de la numérisation des documents a fait naître le besoin d'un système de suivi et de protection, afin de pouvoir contrôler la diffusion de ces documents. La solution étudiée depuis quelques années est le tatouage numérique. Il consiste à insérer au sein du document hôte une marque invisible. Dans l'étude présentée ici, le marquage est robuste (même si le document est modifié, la marque doit pouvoir être extraite) et son extraction doit pouvoir se faire sans l'aide du document original (tatouage aveugle).

Les premières études sur le tatouage étaient basées sur des constatations empiriques. Ce n'est que lorsque l'analogie avec la communication d'un message *via* un canal bruité a été faite que le domaine devint plus formalisé et théorisé. La marque est alors considérée comme un signal que l'on souhaite transmettre sur un canal bruité par le document hôte et les modifications (les attaques) qu'il subit. Ce canal de communication est souvent modélisé par un canal gaussien [10, 7] (canal AWGN [1]).

Alors que le document hôte était inclus dans le bruit pouvant limiter les performances du schéma de tatouage, les études récentes montrent qu'au contraire, en considérant le tatouage comme un problème de communication avec information adjacente [3], le document hôte n'influe pas sur la performance de la transmission. Costa [2] a montré un algorithme théorique permettant d'atteindre la limite de capacité de ce type de canal, appliqué aux signaux i.i.d. gaussiens et à une transmission AWGN. Mais ce schéma est basé sur un dictionnaire de très grande taille de mots de codes aléatoires, non utilisable de façon réaliste. Les approches précédemment proposées [1, 5] pour construire ce dictionnaire ne permettent pas de régler finement les paramètres du code (rendement et taille des sous-dictionnaires). Nous rappelons dans la première partie le principe du schéma de Costa. On introduisons ensuite notre technique de construction d'un dictionnaire structuré, s'appuyant sur des codes correcteurs poinçonnés. Enfin, dans la partie présentant nos résultats, nous appliquons ce type de dictionnaire au sein d'un schéma de tatouage d'images basé sur l'étalement de spectre.

1. Pour *Additive White Gaussian Noise*.

1 Tatouage avec information adjacente

Le tatouage d'un document peut être considéré comme un problème de communication avec information adjacente disponible à l'encodeur [3]. Considérons un signal hôte i.i.d. et gaussien $\vec{x} = \{x_1, x_2, \dots, x_n\}$ modélisé par $X \sim \mathcal{N}(0, Q)$. Ce signal est disponible lors de l'étape de tatouage. Les données sont transmises par l'ajout d'un signal de tatouage $\vec{w} = \{w_1, w_2, \dots, w_n\}$, dont l'énergie est bornée :

$$\frac{1}{n}\sum_{i=1}^{n} w_i^2 \leq P. \tag{1}$$

Le signal marqué est alors $\vec{y} = \vec{x} + \vec{w}$. Durant la transmission, ce signal peut être attaqué. Cette attaque est modélisée par l'ajout d'un bruit gaussien $\vec{z}$, correspondant à $Z \sim \mathcal{N}(0, N)$. Le signal reçu est $\vec{y'} = \vec{y} + \vec{z}$, comme le montre la figure 1.

Si l'on considère le canal comme un simple canal gaussien, le signal transmis $\vec{w}$ est perturbé par $\vec{x}$ et $\vec{z}$. La capacité de ce canal est donc

$$C = \log_2\left[1 + \frac{P}{Q+N}\right]. \tag{2}$$

Or, Costa a montré que le bruit présent et connu lors de l'insertion n'a pas d'influence sur la capacité du canal. Il a conçu

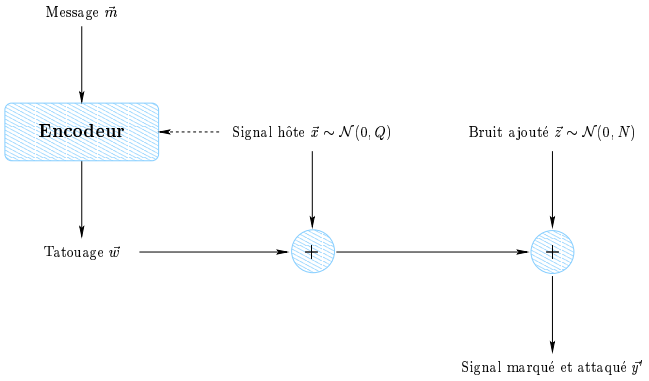


FIG. 1: *Le tatouage vu comme de la communication avec information adjacente*

un algorithme théorique permettant d'atteindre cette capacité, donnée par

$$C = \log_2 \left[1 + \frac{P}{N}\right]. \tag{3}$$

Cette limite est obtenue par l'introduction d'un signal U, obtenu par un dictionnaire structuré : chaque message possible $\vec{m}$ est associé à un sous-dictionnaire $\mathcal{U}_{\vec{m}}$ composé de $2^{nI(U;X)}$ mots de codes[2]. Lors de la phase d'insertion de la marque, le mot de code $\vec{u}^\star \in \mathcal{U}_{\vec{m}}$ le plus proche de $\vec{x}$ est choisi. Le signal de tatouage est alors donné par $\vec{w} = \vec{u}^\star - \alpha\vec{x}$, avec $\alpha = P/(P+N)$. À la réception de l'image, lors de la phase de décodage, le mot de code appartenant à $\mathcal{U}$ le plus proche de $\vec{y'}$ est recherché. Trouver à quel sous-dictionnaire $\mathcal{U}_{\vec{m}}$ il appartient donne le message $\vec{m}$.

Le schéma de Costa utilise pour la construction de son dictionnaire des mots de code pseudo-aléatoires. La recherche du mot de code le plus proche passe donc par une exploration exhaustive du dictionnaire. Elle n'est pas réaliste en pratique. En structurant le dictionnaire et grâce à des algorithmes adaptés, les codes correcteurs d'erreurs (ECC) permettent une recherche rapide de mot de code le plus proche d'un signal donné. Le chapitre suivant expose une technique de création à partir d'un ECC convolutif.

2 Dictionnaire structuré par codes poinçonnés

Les ECC représentent une approche de choix pour construire un dictionnaire simple à explorer. Une des caractéristiques nécessaires au fonctionnement du schéma de Costa est la division du dictionnaire $\mathcal{U}$ en sous-dictionnaires. Les ECC classiques doivent donc être adaptés à cette exigence.

Dans le schéma de Costa, chaque message possible (dont la longueur est notée k ici) est associé à $2^{nI(U;X)}$ mots de codes. Un moyen simple de créer un dictionnaire structuré est d'ajouter au k bits initiaux du message $i = n \times I(U;X)$ bits, qui servent à indexer les différents mots de codes. Ces $k+i$ bits sont codés en utilisant un code correcteur d'erreurs, aboutissant à des mots de code de longueur $(k+i)/r$ où r est le rendement du code. Or, comme i dépend du signal hôte, la longueur des mots de code varie suivant le signal hôte. Potentiellement, à des signaux hôtes de même dimension peuvent donc correspondre des dictionnaires dont les mots de code n'ont pas la même longueur. Cette constatation, que l'on également faire sur les codes de Chou *et al.* [1] ou de Miller [5], rend ce schéma impossible à utiliser tel quel. La dimension du signal hôte est en effet fixe (dimension d'une images, taille d'un échantillon sonore, ...) et les mots de code doivent être de même dimension pour qu'une insertion de type $\vec{y} = \vec{x} + \vec{w}$ soit possible. Le rendement global k/n doit donc être fixe et indépendant de i.

Nous proposons donc d'utiliser un dictionnaire basé sur un code convolutif poinçonné et sur un décodage par treillis souple. Considérons un code correcteur de rendement $r = k/n$. Nous construisons un motif de taille $k+i$ composé des k bits du message à transmettre $\vec{m}$ et de i bits supplémentaires, dont les valeurs ne sont pas fixés et qui serviront à indexer les différent mots de code du sous-dictionnaire $\mathcal{U}_{\vec{m}}$ (voir la figure 2(a)). Ce motif est entrelacé afin d'avoir une bonne répartition des mots de code. Le signal hôte $\vec{x}$ de taille initiale n est étendu à $(k+i)/r$ valeurs par insertion de valeurs neutres (c'est-à-dire 0). Ce signal hôte étendu est décodé en utilisant un algorithme de Viterbi modifié : le motif construit précédemment est utilisé comme fort *a priori* afin que les k bits du message à transmettre forcent certaines transitions au sein du treillis convolutif (figure 2(b)). Le mot de code obtenu correspond bien au message $\vec{m}$ grâce aux k bits fixés du motif. Il est poinçonné en accord avec l'extension précédente du signal hôte, aboutissant à un mot de code de dimension n et répondant aux impératifs du schéma de Costa (mot de code $\vec{u}^\star \in \mathcal{U}_{\vec{m}}$ et plus proche de $\vec{x}$).

À partir de $\vec{u}^\star$, la marque est choisie en maximisant la robustesse, tel que décris par Miller *et al.* [6]. Dans l'espace à n dimensions, à $\vec{u}^\star$ correspond un hyper-cône de robustesse où $\vec{y'}$ doit se trouver pour que le décodage se fasse correctement. La marque $\vec{w}$ est alors définie afin que $\vec{x}+\vec{w}$ se trouve à l'intérieur de ce cône, le plus loin possible des bords (afin de maximiser la robustesse, c'est-à-dire l'énergie du bruit d'attaque qu'il faudra ajouter pour sortir de ce cône) :

$$\vec{w} = \arg\max_{\widetilde{\vec{w}}} \left\{ R(\vec{u}^\star, \vec{x} + \widetilde{\vec{w}}) \right\} \tag{4}$$

avec

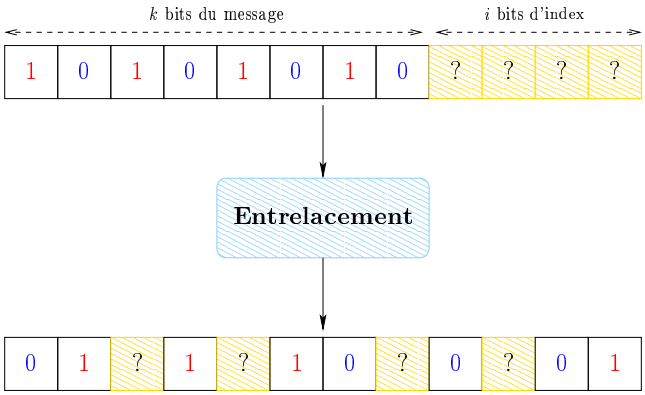
$$R(\vec{u}, \vec{y}) = \left[\frac{\vec{y}\cdot\vec{u}}{\|\vec{u}\|}\right]^2 \left(1 + \tan^2\theta\right) - \|\vec{y}\|^2, \tag{5}$$

et où θ est l'angle de l'hyper-cone, donné par [8]

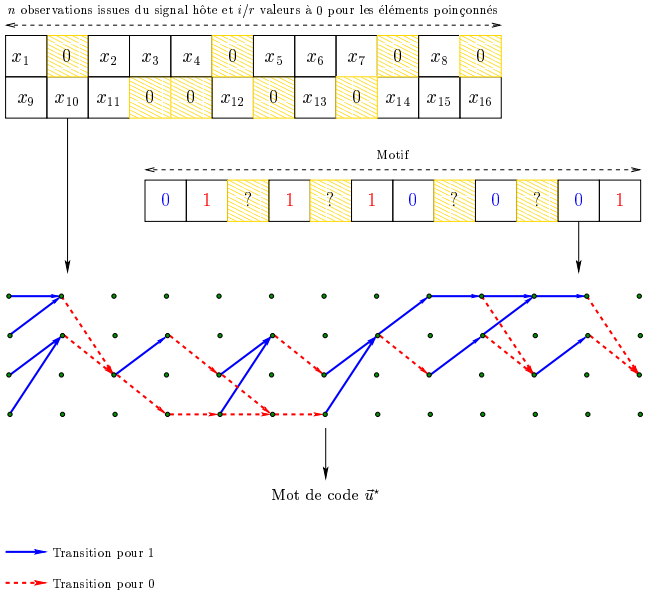
$$\tan^{-2}\theta = 2^{\frac{2(k+i)}{n}} - 1. \tag{6}$$

À l'extraction, le signal $\vec{y'} = \vec{y} + \vec{z}$ est étendu (comme lors du décodage de $\vec{x}$ pendant l'insertion) et décodé en utilisant l'algorithme de Viterbi appliqué au treillis convolutif complet. Grâce au décodage souple et au fait que tous les mots de code soient de même énergie, ce schéma de codage est insensible aux facteurs d'échelle : $\vec{y'}$ peut être multiplié par un scalaire sans impact sur la robustesse du décodage, avantage important dans le cas d'attaques de type SAWGN [11, 4].

2. Où $I(U;X)$ est l'information mutuelle entre U et X.



(a) Construction d'un motif pour le décodage de $\vec{x}$



(b) Décodage du signal hôte grâce à un algorithme de Viterbi avec un treillis dont seuls les transitions compatibles avec le motif ont été conservées

FIG. 2: *Recherche du mot de code le plus proche lors de la phase d'insertion ($k = 8, i = 4, n = 16$ et $r = 1/2$)*

3 Résultats : application à l'image

Afin de mesurer la distorsion introduite par la phase d'insertion de la marque et par les attaques, nous utilisons une erreur quadratique moyenne pondérée, définie par

$$D_{xy} = \frac{1}{m} \sum_{i=1}^{m} \varphi_i^2 \left(x_i - y_i\right)^2, \tag{7}$$

où φ_i est une pondération perceptuelle et m la taille de l'image. Dans les résultats de cette section, qui présente notre méthode appliquée aux images en niveaux de gris, cette mesure est inspirée de celle de Watson [12] et est définie par

$$\varphi_i = \frac{\rho}{\sqrt{a(X_i) + 1}}, \tag{8}$$

avec ρ fixé afin que la moyenne des φ_i sur l'image considérée soit de 1, et $a(X_i)$ la mesure d'activité locale du $i^{\text{ème}}$ élément du signal hôte (basée sur sa variance locale).

Afin d'obtenir le signal hôte $\vec{x}$, l'image de test est transformée en ondelettes sur trois niveaux. En utilisant une fonction d'étalement de spectre[3], nous définissons un sous-espace i.i.d. gaussien de taille $n = 132$, définissant ainsi le signal hôte $\vec{x}$. Ce signal est marqué à partir d'un message de $k = 64$ bits en fixant une distorsion d'insertion $D_{xy} = 7$ (très bonne qualité visuelle, comme le montre la figure 3). Le message est codé en utilisant la technique présentée ici, avec un rendement $r = 1/2$. L'image marquée est ensuite attaquée.

Pour chaque niveau d'attaque testé, la distorsion introduite $D_{xy'}$ (distorsion entre l'image originale et l'image marquée et attaquée) est calculée, et le message est extrait. La performance du schéma est mesurée par le rapport signal à bruit, noté E_b/N_0, du message extrait.

La figure 4(a) montre le comportement du schéma face à l'ajout de bruit gaussien de plus en plus fort. L'abscisse représente la distorsion introduite par l'attaque et l'ordonnée mesure la performance du schéma (rapport signal à bruit du canal de tatouage). De même, la figure 4(b) mesure la résistance face à la compression de type JPEG, de qualité de plus en plus faible. Pour permettre un comparaison, les performances d'un schéma similaire, mais sans prise en compte de l'information adjacente, sont indiquées en pointillés.

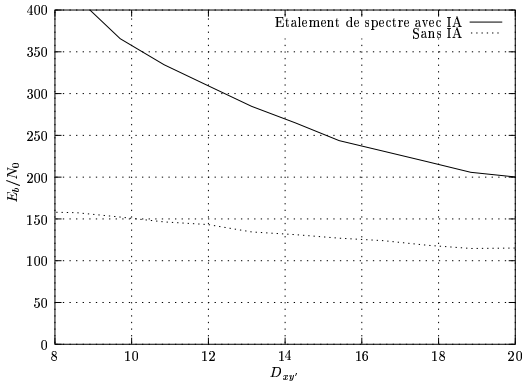
Les résultats montrent dans les deux cas une très forte supériorité du schéma prenant en compte l'information adjacente et utilisant notre dictionnaire structuré.

4 Conclusion

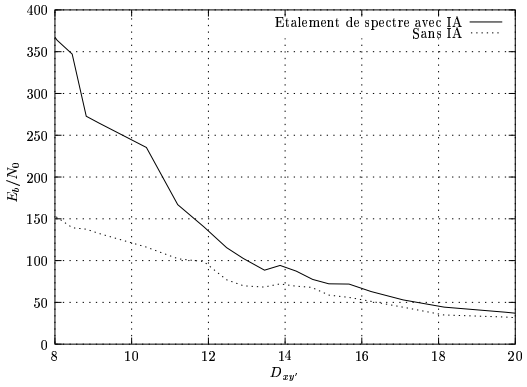
Nous avons étudié dans ce papier la construction d'un dictionnaire structuré, adapté au tatouage avec prise en compte de l'information adjacente présente à l'encodage. Cela est rendu possible par des codes correcteurs d'erreurs convolutifs poinçonnés. Les phases d'encodage et de décodage se basent sur l'algorithme de Viterbi, assurant ainsi une faible complexité. Les tests sur des images en niveaux de gris démontrent la supériorité de ce type de schéma par rapport au simple tatouage additif.

3. Projection des données issues de la transformée en ondelettes sur des porteuses pseudo-aléatoires [8].

FIG. 3: *Image Paper machine* 512×512 *marquée avec une distorsion d'insertion* $D_{xy} = 7$ *(copyright photo courtesy of Karel de Gendre, extrait de la base de photos conseillées pour Stirmark [9])*



(a) Performance contre l'ajout de bruit gaussien.



(b) Performance contre la compressions JPEG ($D_{xy'} = 7$ pour une qualité de 100 % et $D_{xy'} \simeq 20$ pour une qualité de 15 %).

FIG. 4: *Rapport signal à bruit pour l'image Paper machine (niveaux de gris, taille* 512×512*,* $k = 64$ *bits,* $n = 132$*, DWT sur 3 niveaux et* $D_{xy} = 7$*).*

Références

[1] J. Chou, S. S. Pradhan, and K. Ramchandran. Turbo coded treillis-based constructions for data embedding: channel coding with side information. In *Proc. Conf. on Signals, System and Computers*, Asilomar, CA, Nov. 2001.

[2] M. H. M. Costa. Writing on dirty paper. *IEEE Trans. on Info. Theory*, 29(3):439–441, May 1983.

[3] I. J. Cox, M. L. Miller, and A. L. McKellips. Watermarking as communications with side information. *Proc. IEEE*, 87(7):1127–1141, Jul. 1999.

[4] J. J. Eggers, R. Bäuml, and B. Girod. Digital watermarking facing attacks by amplitude scaling and additive white noise. In *4th Int. ITG Conf. on Source and Channel Coding*, Jan. 2002.

[5] M. Miller, G. J. Doërr, and I. J. Cox. Dirty-paper trellis codes for watermarking. In *Proc. Int. Conf. on Image Processing*, Rochester, NY, Sep. 2002.

[6] M. L. Miller, I. J. Cox, and J. A. Bloom. Informed embedding: exploiting image and detector information during watermark insertion. In *Proc. Int. Conf. on Image Processing*, Vancouver, Canada, Sep. 2000.

[7] P. Moulin and J. A. O'Sullivan. Information-theoretic analysis of information hiding. *IEEE Trans. on Info. Theory*, Oct. 1999.

[8] S. Pateux and G. Le Guelvouit. Practical watermarking scheme based on wide spread spectrum and game theory. *IEEE Trans. on Image Communication*, (18), 2003.

[9] F. A. P. Petitcolas, R. J. Anderson, and M. G. Kuhn. Attacks on copyright marking systems. In *Proc. Int. Workshop on Information Hiding*, pages 219–239, Portland, USA, Apr. 1998.

[10] S. D. Servetto, C. I. Podilchuk, and K. Ramchandran. Capacity issues in digital image watermarking. In *Proc. Int. Conf. on Image Processing*, volume 1, pages 445–449, Chicago, IL, Oct. 1998.

[11] 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*, 81(6), Jun. 2001.

[12] A. B. Watson. DCT quantization matrices visually optimized for individual images. *Proc. SPIE*, 1913:202–216, 1993.