Contribution des filtres LPTV et des techniques d'interpolation au tatouage numérique.

Vincent Martin

N° d'ordre :

THÈSE

présentée pour obtenir le titre de

DOCTEUR DE L'INSTITUT NATIONAL POLYTECHNIQUE DE TOULOUSE

École doctorale : Informatique et Télécommunications
Spécialité : Signal, Image, Acoustique et Optimisation

par

Vincent Martin

CONTRIBUTION DES FILTRES LPTV ET DES TECHNIQUES D'INTERPOLATION AU TATOUAGE NUMÉRIQUE

soutenue le 28 novembre 2006 devant le jury composé de

M. Jean-Yves TOURNERET	Professeur à l'E.N.S.E.E.I.H.T., Toulouse	Président
M. Pierre DUHAMEL	Directeur de Recherche C.N.R.S. au L.S.S., Gif sur Yvette	Rapporteur
M. Benoît MACQ	Professeur à l'Université catholique de Louvain	Rapporteur
M. Jordi INGLADA	Ingénieur de Recherche C.N.E.S., Toulouse	Membre
Mme. Marie CHABERT (*)	Maître de Conférence à l'E.N.S.E.E.I.H.T., Toulouse	Membre
M. Bernard LACAZE (*)	Professeur à l'I.N.S.A., Toulouse	Examinateur

(*) Directeur de thèse

Résumé

Les Changements d'Horloge Périodiques (PCC) et les filtres Linéaires Variant Périodiquement dans le Temps (LPTV) sont utilisés dans le domaine des télécommunications multi-utilisateurs. Dans cette thèse, nous montrons que, dans l'ensemble des techniques de tatouage par étalement de spectre, ils peuvent se substituer à la modulation par code pseudo-aléatoire. Les modules de décodage optimal, de resynchronisation, de pré-annulation des interférences et de quantification de la transformée d'étalement s'appliquent également aux PCC et aux filtres LPTV. Pour le modèle de signaux stationnaires blancs gaussiens, ces techniques présentent des performances identiques à l'étalement à Séquence Directe (DS) classique. Cependant, nous montrons que, dans le cas d'un signal corrélé localement, la luminance d'une image naturelle notamment, la périodicité des PCC et des filtres LPTV associée à un parcours d'image de type Peano-Hilbert conduit à de meilleures performances. Les filtres LPTV sont en outre un outil plus puissant qu'une simple modulation DS. Nous les utilisons pour effectuer un masquage spectral simultanément à l'étalement, ainsi qu'un rejet des interférences de l'image dans le domaine spectral. Cette dernière technique possède de très bonnes performances au décodage.

Le second axe de cette thèse est l'étude des liens entre interpolation et tatouage numérique. Nous soulignons d'abord le rôle de l'interpolation dans les attaques sur la robustesse du tatouage. Nous construisons ensuite des techniques de tatouage bénéficiant des propriétés perceptuelles de l'interpolation. La première consiste en des masques perceptuels utilisant le bruit d'interpolation. Dans la seconde, un schéma de tatouage informé est construit autour de l'interpolation. Cet algorithme, qu'on peut relier aux techniques de catégorisation aléatoire, utilise des règles d'insertion et de décodage originales, incluant un masquage perceptuel intrinsèque. Outre ces bonnes propriétés perceptuelles, il présente un rejet des interférences de l'hôte et une robustesse à diverses attaques telles que les transformations valumétriques. Son niveau de sécurité est évalué à l'aide d'algorithmes d'attaque pratiques.

Summary

Periodic Clock Changes (PCC) and Linear Periodically Time Varying (LPTV) filters have previously been applied to multi-user telecommunications in the Signal&Communications group of IRIT laboratory. In this thesis, we show that in each digital watermarking scheme involving spread-spectrum, they can be substituted to modulation by a pseudo-noise. The additional steps of optimal decoding, resynchronization, pre-cancellation of interference and quantization of a spread transform apply also to PCCs and LPTV filters. For white Gaussian stationary signals, these techniques offer similar performance as classical Direct Sequence (DS) spreading. However we show that, in the case of locally correlated signals such as image luminance, the periodicity of PCCs and LPTV filters associated to a Peano-Hilbert scan leads to better performance.

Moreover, LPTV filters are a more powerful tool than simple DS modulation. We use LPTV filters to conduct spectrum masking simultaneous to spreading, as well as image interference cancellation in the spectral domain. The latter technique offers good decoding performance.

The second axis of this thesis is the study of the links between interpolation and digital watermarking. We stress the role of interpolation in attacks on the watermark. We propose then watermarking techniques that benefit from interpolation perceptual properties. The first technique consists in constructing perceptual masks proportional to an interpolation error. In the second technique, an informed watermarking scheme derives form interpolation. This scheme exhibits good perceptual properties, host-interference rejection and robustness to various attacks such as valumetric transforms. Its security level is assessed by *ad hoc* practical attack algorithms.

Remerciements

Je voudrais remercier chaleureusement les personnes qui m'ont accompagné durant ces trois années de thèse, et qui les ont rendues si appréciables.

Je remercie avant tout mes directeurs de thèse. Merci à Marie Chabert pour la confiance qu'elle m'accordée et la gentillesse, le soin et la disponibilité avec lesquels elle a accompagné mes travaux. Merci à Bernard Lacaze pour ses conseils éclairants.

Je tiens bien sûr à remercier Pierre Duhamel et Benoît Macq pour avoir accepté d'être rapporteurs de cette thèse (de ce pavé ?). Je remercie également Jordi Inglada d'avoir voulu s'intéresser à mes travaux. Merci enfin à Jean-Yves Tourneret d'avoir présidé ce jury avec la sobriété et la justesse d'un maître florentin. Merci à tous d'avoir rendu la soutenance agréable et constructive.

La vraie réussite de cette thèse, ce sont les amitiés qu'elle m'a permis de forger. Merci donc à tous ceux qui ont partagé mon bureau, ou tout comme : Ana la kidnappée de l'Onera, Virginie (potins de premier choix), Olivier le gentleman du XIII, Audrey "ho mais c'est horrible", Vincent le conteur venu du Sud, Florent (merci pour tes interprétations de la Comédie Musicale de la Révolution, pour tes actes politiques majeurs anti-berlusconiens et pour le parcours de Peano), Mathieu (est-il mon idole ?), Jean-Pierre (PJ c'est plus sexy) et Farid l'homme-puma, sans oublier les petits nouveaux Jean-Rémi et Frédéric. Spéciale dédicace au maître Wilfried Chauvet. Non content d'avoir dépoussiéré les LPTV (il m'a beaucoup aidé !), il est la mémoire vivante du TéSA et de l'IRIT. Je me souviendrai également de la joyeuse présence de Milena, David, Garmy, Fred, Sakuna, Nicolas, Ridha, Alexandra, Rahim, Farf'... Cette thèse a vu apparaître la tradition de "l'Anglaise", ayons donc une pensée pour nos coéquipiers du PL : Cyrielle, Valérie, Antoine, Denis, Aniela, Hélène et Fred. Merci à Sylvie pour l'aide précieuse qu'elle apporte à tous ses petits doctorants. Merci également à tous les autres membres du labo pour la sympathie qu'il m'ont témoigné : Nathalie, Corinne, Martial, Jérôme, André-Luc, Benoît, Marie-Laure, Patrice, Marc, Jean-Luc, Manu, Daniel, Julien, Xavier... Je garderai enfin un chouette souvenir des équipées à Barcelone, Antalya, Caen ou Florence, et de ceux qui y ont participé.

Merci à ma tribu paléo-toulousaine, en espérant continuer à construire à vos côtés : Thierry, Chloé, Serge, Elodie, Arnaud, Marie, Mehdi, Camille, Julien, Anne-Laure, Rémi, Emma, Delphine, Nicolas, Caroline, Jean-Marc, Cécile, Ségolène, Raphaël, Séverine, Yannick, Matthieu, Jérôme et Simon. Bienvenue à Arthur, Camille et Lucas !

Un grand merci à mes parents pour leur soutien constant et la stabilité qu'ils m'ont apportée. Merci également pour le succès du pot ! Merci à Aurélie pour sa présence, ainsi qu'à Laurent pour son enthousiasme communicatif. Merci à Pauline, Elisa, Marie-Christine et Paul, vous faites désormais partie de ma famille.

Le temps et le soin que j'ai consacrés à cette thèse sont dédiés à Juliette. Merci pour ton aide, pour ton courage dans les épreuves que nous avons surmontées et pour les moments de bonheur que nous continuons à partager.

Notations

Glossaire

Méthodes de tatouage proposées

1D-PCC, 2D-PCC : tatouage par PCC uni- ou bidimensionnels
PCC+Peano : version de 1D-PCC combinée à un parcours de Peano-Hilbert
LL-LPTV : tatouage par filtres LPTV sans perte
mod-LPTV : tatouage par filtres LPTV à filtres modulateurs constants
orth-LPTV : version de mod-NRZ-LPTV avec orthogonalité entre utilisateurs
ZI-LPTV : tatouage par filtres LPTV à insertion de zéros
type-NRZ-LPTV : variante des méthodes précédentes utilisant la mise en forme NRZ
mask-LPTV : tatouage par filtres LPTV opérant un masque spectral
W-interp : méthode de tatouage utilisant l'interpolation
W-bilin, W-spline : implantations de W-interp utilisant l'interpolation bilinéaire ou B-spline bicubique
DC-W-interp : extension de W-interp à la compensation des distorsions

Méthodes de tatouage classiques

L-*méthode* : insertion dans le domaine spatial (luminance)
DCT-*méthode* : insertion dans le domaine de la DCT par blocs 8x8
méthode+W : variante utilisant un préfiltrage de Wiener au décodage
DS : méthode de tatouage utilisant les Séquences Directes
NVF : masque psychovisuel de Fonction de Visibilité du Bruit
ISS : étalement de spectre amélioré
LISS : étalement de spectre amélioré linéaire
QIM : Modulation d'Indices de Quantification
DM, STDM : QIM à signal d'agitation et sa version à Transformation d'Étalement
SCS, ST-SCS : Schéma de Costa Scalaire et sa version à Transformation d'Étalement
QP : méthode de Projection Quantifiée
SSP : quantification de la Projection sur un Sous-Espace
RDM : DM sur une composante Rationnelle du signal

Abréviations

PCC : Changement d'Horloge Périodique
LPTV : Filtre Linéaire Périodique Variant dans le Temps
DCT : Transformée en Cosinus Discrète bidimensionnelle par blocs 8x8
TEB : Taux d'Erreur Binaire (critère de performance au décodage)

COR : Caractéristiques Opérationnelles du Récepteur (critère de détection)
DWR : Rapport Document à Tatouage
WNR : Rapport Tatouage à Bruit
DNR : Rapport Document à Bruit
SNR : Rapport Signal à Bruit (en télécoms traditionnelles)
PSNR, PSNR$_B$: Rapport entre le Pic du Signal et le Bruit (critère psychovisuel), pour le tatouage ou pour l'attaque
EQM : Erreur Quadratique Moyenne
NC : Corrélation Normalisée (critère de détection)
MAI : Interférences d'Accès Multiple (entre utilisateurs)
GGD : Distribution Gaussienne Généralisée
JND : Différence Juste Perceptible
NRZ : Non Retour à Zéro (mise en forme)
FIM : Matrice d'Information de Fisher

Notations

Notations mathématiques

X : variable aléatoire
$\mathbf{x}$: selon le contexte, vecteur ou matrice
x_k : élément k de X pris comme un vecteur
x_{k_1,k_2} : élément (k_1, k_2) de X pris comme une matrice
μ_X,σ_X^2 : moyenne et variance de la variable aléatoire X
t_1, t_2 : coordonnées horizontale et verticale continues
$f(.)$,f^{-1} : fonction et son inverse éventuel
$*$: opération de convolution
$\times$: produit cartésien
$<,>$: produit scalaire
δ : symbole de Kronecker
Int : partie entière
$\lfloor n \rfloor, \lceil n \rceil$: reste et quotient de la division euclidienne de n par T
$\mathcal{N}(\mu, \sigma)$: loi normale de moyenne μ et de variance σ
β, c : paramètres d'une GGD
Q : fonction d'erreur gaussienne
P_d : probabilité de détection
P_{nd}, P_{fa}, P_e : probabilités de non-détection, de fausse alarme et d'erreur
H_0, H_1, H_{-1} : hypothèses d'un test statistique
η : seuil de décision d'un test statistique
$\mathbf{T}$: statistique de test
p$(a|b)$: probabilité (que $A = a$ sachant que $B = b$)
F$_X$, f_X : fonction de répartition et densité de probabilité de la variable aléatoire X
K_X : fonction d'autocorrélation de X
$E[X]$: espérance de X
$\mathcal{P}_X$: puissance de X
I$(X;Y)$: information mutuelle de X et Y
$H(X)$: entropie de X
∇_X : gradient de X
$\|X\|$: norme de X

$X(\omega)$, $X(z)$: transformée de Fourier et transformée en Z de $\mathbf{x}$

Notations liées au tatouage

Notations générales

$\mathbf{x}$: document original
$\mathbf{y}$: document tatoué
$\mathbf{z}$: document attaqué
$\mathbf{z}'$: préfiltrage de Wiener de $\mathbf{z}$
$\mathbf{w}$: tatouage
$\mathbf{w}'$: tatouage préfiltré (ISS)
$\mathbf{n}$: bruit additif
$\mathbf{m}$: message transmis par le tatouage (éléments : m_l)
$\mathbf{b}$: message mis en forme
$\mathbf{k}$: clé secrète
$\mathbf{c}$: séquence pseudo-aléatoire (code)
$\mathbf{x^j}$: vecteur $\mathbf{x}$ correspondant à l'utilisateur j
$\hat{\mathbf{x}}$: estimation de $\mathbf{x}$ (par exemple par un filtre moyenneur)
$\hat{\mathbf{m}}$: décision finale sur le message décodé
$\mathbf{d}$: estimation de $\mathbf{m}$ au décodage
$\underline{\mathbf{x_l}}$: sous-vecteur de $\mathbf{x}$ de taille P et de coordonnées $\in \mathcal{S}_l$
$t_{\mathbf{x}}(u,v)$: DCT 2D de $\mathbf{x}$
$t^k_{\mathbf{x}}(u,v)$: coefficient de coordonnées (u,v) du bloc k de la DCT 2D par blocs 8x8
N_1, N_2 nombre de lignes et de colonnes de $\mathbf{x}$
N : taille de $\mathbf{x}$, si elle est considérée comme un vecteur ; $N = N_1 N_2$
L : taille de $\mathbf{m}$ (nombre de bits d'information insérés ou charge utile)
P : redondance à l'insertion ; $P = N/L$
J : nombre d'utilisateurs
Ψ : masque psychovisuel
h_ψ : réponse impulsionnelle de certains masques psychovisuels
γ : facteur de correction gamma
θ : angle d'une rotation
$x_{\max}$: valeur maximale qu'un point de $\mathbf{x}$ peut prendre
$\mathcal{S}$: ensemble d'insertion
$\mathcal{S}_l$: ensemble d'insertion associé au bit l
$\mathcal{U}$: dictionnaire de mots de code
$\mathcal{M}$: dictionnaire des mots de codes (tatouages)
$\mathcal{M}_{\mathbf{m}}$: sous-dictionnaire correspondant au message $\mathbf{m}$
Sp : étape d'étalement dans un algorithme de tatouage
Tr : transformation inversible utilisée dans un algorithme de tatouage
$\mathcal{A}$: attaque
$g(.)$: fonction d'insertion
$c(.,.)$: contrainte
D_W : distance de Watson
D_{KL} : distance de Kullback-Leibler
N_o : nombre d'observations
N_c : nombre de clés
N_o^* : borne de sécurité sur N_o

Notations spécifiques au tatouage informé

λ : coefficient de compensation des distorsions dans LISS
α : coefficient de compensation des distorsions dans SCS ; coefficient de modulation du message dans LISS
Δ : pas de quantification
$Q_{\Delta,\tau}$: quantificateur de pas Δ et de décalage initial τ
$\mathbf{u}$: signal auxiliaire du schéma de Costa
D_w : distorsion réelle (*a posteriori*)
q : erreur de quantification
P_s : taille du code d'étalement de QP

Notations spécifiques aux filtres LPTV

T, T_{1D}, T_{2D} : période d'un PCC ou d'un filtre LPTV (1D ou 2D)
f : fonction PCC
q : permutation aléatoire
$\mathcal{F}^{\text{LPTV}}$: filtre LPTV
$\mathbf{u}$: entrée du filtre
$\mathbf{v}$: sortie du filtre
A : matrice de filtrage dans LL-LPTV, mod-LPTV, ZI-LPTV, mask-LPTV
$h(n)$: réponse impulsionnelle du filtre
ν : interférences de l'hôte (IPCC)
$\mathbf{q}$: mise en forme de l'erreur de quantification (LPTV-SCS)

Notations spécifiques à l'interpolation

$\hat{x}(t), \hat{x}(t_1, t_2)$: interpolation à partir de $\mathbf{x}$ au point t ou (t_1, t_2)
$x(t)$: signal continu dont $\mathbf{x}$ est un échantillonnage
f^n : fonction de synthèse d'ordre n
$\eta^n(.)$: spline cardinale d'ordre n
$\beta^n(.)$: B-spline d'ordre n
$\mathcal{G}$: grille d'interpolation
$\mathbf{g} = \{g^k\}$: ensemble des fonctions interpolantes
g_j^k : poids d'un pixel dans l'interpolation par g^k
N_v : support de la fonction interpolante
$N_{\mathcal{S}}$: nombre de points de $\mathbf{x}$ substitués dans W-interp
$P_{\mathcal{S}}$: redondance à l'insertion de W-interp ; $P_{\mathcal{S}} = N_{\mathcal{S}}/L$
τ^u, τ^v : décalages aléatoires utilisés dans W-interp
$\mathcal{T}$: ensemble des décalages aléatoires
η_{th} : seuil de décision théorique pour W-interp
$\mathbf{r}$: comparaison au décodage
$\epsilon(\mathbf{x})$, $\epsilon(\mathbf{n})$: contributions du document tatoué et du bruit à $\mathbf{r}$
ρ_l : erreur au sens de l'EQM sur $\mathcal{S}_l$
Δ : pondération de la puissance du bruit selon $\mathcal{T}$

Introduction

Le tatouage numérique est un thème de recherche nouveau dans l'équipe Signal et Communications (SC) du laboratoire IRIT à Toulouse. Cependant, ses liens avec les techniques de télécommunications ont conduit M. Chabert et B. Lacaze à proposer ce sujet de thèse en juin 2003. En particulier, le point de départ de ce travail a été l'application au tatouage numérique des Changements d'Horloge Périodiques, technique initialement proposée par B. Lacaze et appliquée par la suite à l'étalement de spectre. Cette étude s'est ensuite orientée vers le cadre plus général et plus puissant des filtres Linéaire Variant Périodiquement dans le Temps, qui avaient eux aussi fait l'objet d'une thèse dans l'équipe SC. Bien que les techniques étudiées puissent s'appliquer à tout type de signal, nous nous sommes concentré sur une application aux images naturelles. D'autre part, à mesure de l'assimilation des principes du tatouage numérique, il nous est apparu que la problématique de l'interpolation y était très liée. Pourtant, son étude est souvent occultée et reléguée au rang d'attaque. Nous avons donc cherché à étudier précisément l'impact de l'interpolation lorsqu'elle sert d'attaque, puis à utiliser ses propriétés perceptuelles au bénéfice de nouveaux algorithmes de tatouage numérique.

Le premier chapitre de cette thèse vise à présenter le contexte général du tatouage numérique. Après une présentation du principe du tatouage et de ses applications, en particulier pour la gestion des droits d'auteur, nous analysons les principes de conception d'une technique de tatouage. On y distingue les techniques de tatouage additif (le plus souvent par étalement de spectre), les techniques substitutives avec dictionnaire (souvent rassemblées sous le terme de "catégorisation aléatoire") et les techniques substitutives avec contraintes, qui seront toutes utiles dans les chapitre suivants. Nous soulignons également le caractère multidisciplinaire du tatouage numérique. Les principales attaques auxquelles le document tatoué peut être soumis sont répertoriées. Les principes de sécurité d'un algorithme de tatouage sont exposés. Enfin, nous présentons les spécificités de l'application aux images naturelles, au travers des domaines transformés et modèles statistiques ou perceptuels utilisables.

Le second chapitre rassemble l'essentiel de nos contributions au tatouage par étalement de spectre. Dans un premier temps, la technique des Changements d'Horloge Périodiques (PCC) est détaillée. Nous proposons deux techniques de tatouage, fondées sur les PCC unidimensionnels (1D-PCC) et bidimensionnels (2D-PCC). Nous y montrons qu'il s'agit d'un modèle très simple d'étalement de spectre multi-utilisateurs, dont les performances théoriques et expérimentales sont globalement similaires à celles des techniques classiques de modulation par un code (DS). Un point essentiel apparaît cependant dans l'étude expérimentale de la robustesse. En effet, il est connu qu'une image naturelle est un support non stationnaire, et il est classique de contourner cette propriété. Nous montrons qu'il est pourtant possible de tirer profit de la corrélation

entre les blocs de pixels de l'image en utilisant une mise en forme particulière et en exploitant la périodicité des PCC. Cette observation conduit à introduire une technique de parcours d'image destinée à préserver la corrélation entre des pixels voisins. L'emploi de parcours tels que celui de Peano-Hilbert est inusité en tatouage, où le message est souvent mis en forme par une dispersion aléatoire sur les deux dimensions de l'image. La combinaison du parcours de Peano-Hilbert avec 1D-PCC dans le domaine spatial (PCC+Peano) apporte une amélioration significative de la robustesse au bruit de l'image hôte par rapport à la technique DS classique.

Les PCC sont un cas particulier de filtres Linéaires Variant Périodiquement dans le Temps (LPTV). Grâce à la périodicité, cet ensemble de filtres offre également une bonne robustesse à l'image hôte lorsqu'il est combiné au parcours de Peano-Hilbert. Nous appliquons tout d'abord au tatouage d'images une technique inspirée des télécommunications : les filtres LPTV sans perte (LL-LPTV). Nous nous inspirons également de travaux sur l'inversibilité des filtres LPTV pour proposer une technique d'étalement à partir des filtres modulateurs (mod-LPTV), qui est étendue au tatouage multiple (orth-LPTV). Ces deux techniques présentent des performances similaires à celles de PCC+Peano. A partir de leur décomposition en filtre modulateurs, les filtres LPTV permettent de surcroît d'imposer des contraintes sur le spectre du signal en sortie, lorsque le spectre de l'entrée est connu. Deux techniques dédiées au tatouage numérique sont donc proposées. La première, appelée filtres LPTV à insertion de zéros (ZI-LPTV) a pour but d'annuler les interférences du document hôte à la réception, en introduisant une forme de modulation particulière du message. Cette démarche visant, durant la phase d'étalement, à tirer profit des informations sur le spectre du document hôte, est originale. C'est une sorte de "tatouage à spectre de l'hôte connu". La seconde technique impose des contraintes sur le spectre du tatouage simultanément à l'étalement, afin de respecter la "Contrainte du Spectre de Puissance". Cette méthode de tatouage, appelée mask-LPTV, a surtout pour but d'illustrer les potentialités des filtres LPTV par rapport à une modulation simple.

Pour l'ensemble des techniques de tatouage par étalement de spectre proposées, nous étendons ensuite la chaîne de tatouage à l'utilisation des propriétés statistiques de l'image au décodage. Les principes de pré-blanchiment avant décodage dans le domaine spatial et de décodage optimal dans le domaine transformé s'appliquent aux algorithmes proposés, qui perdent cependant une partie de leurs spécificités. Nous adaptons ensuite la chaîne de tatouage proposée dans ce chapitre au tatouage informé. L'exploitation de la connaissance du document à l'insertion permet d'éliminer les interférences qu'elle peut créer à la réception. L'étalement par filtres LPTV peut se substituer dans de nombreux cas à l'étalement par code DS. Nous proposons donc les PCC améliorés (IPCC) et les filtres LPTV améliorés (ILPTV), qui utilisent une pré-annulation des interférence de l'hôte. L'étalement de spectre est également présent dans les techniques de quantification à redondance par étalement. Pour l'ensemble de ces techniques, nous montrons que l'étalement par filtre LPTV peut être utilisé comme technique de projection (PCC-SCS ou LPTV-SCS). Enfin, nous calculons le niveau de sécurité théorique des techniques proposées par application des résultats connus pour la technique DS. Nous proposons en outre un algorithme pratique d'attaque sur la sécurité des PCC, utilisant une estimation itérative de la clé.

Le troisième chapitre de cette thèse permet d'effectuer une transition vers la problématique de l'interpolation. En effet, de nombreuses attaques telles que les attaques géométriques génèrent un bruit d'interpolation, à cause du rééchantillonnage. Même si une resynchronisation est possible, ce bruit ne doit pas être négligé. Tout d'abord, nous

étudions la robustesse de techniques d'étalement de spectre fondées sur les PCC et les filtres LPTV à ce bruit d'interpolation. Si le bruit d'interpolation est fréquent dans les attaques sur le tatouage, c'est qu'il possède d'excellentes propriétés perceptuelles. La démarche adoptée dans un second temps consiste à exploiter les propriétés perceptuelles de l'interpolation au bénéfice du tatouage numérique. Une classe de masques perceptuels fondés sur l'erreur d'interpolation est donc proposée.

Enfin, dans le quatrième chapitre, nous poussons plus loin ce principe pour proposer une classe de schémas de tatouage construits autour de l'interpolation (W-interp). Les diverses contraintes du problème conduisent à utiliser une technique de décodage proche des techniques de catégorisation aléatoire. Comme ces dernières, la technique proposée bénéficie d'un rejet des interférences de l'hôte. Des stratégies d'insertion informée sont également applicables. L'utilisation de l'erreur d'interpolation fournit de surcroît de bonnes propriétés perceptuelles, et une meilleure robustesse à certaines attaques comme les attaques valumétriques. L'application de W-interp au tatouage d'images naturelles est plus particulièrement étudiée. Deux cas particuliers de techniques d'interpolation sont utilisés : l'interpolation bilinéaire (W-bilin) et l'interpolation par les B-splines cubiques (W-spline). On montre que l'erreur d'interpolation d'une image naturelle suit une distribution gaussienne généralisée, ce qui permet de construire un décodeur optimal pour cette distribution. Le niveau de sécurité de W-interp est finalement étudié. Des algorithmes pratiques d'attaques sur la sécurité de W-interp utilisant un algorithme d'Espérance-Maximisation sont proposés.

La lecture du chapitre 1, qui est bibliographique, ne sera pas indispensable au lecteur familier des problématiques du tatouage. La lecture de la partie 3.1.1 ne sera pas indispensable au lecteur familier des techniques d'interpolation. D'autre part, les chapitres 2 et 4 sont indépendants, le chapitre 3 faisant la transition entre les deux approches.

Table des matières

Chapitre 1

Introduction au tatouage d'images numériques

Sommaire

Ce chapitre présente le principe du tatouage de documents numériques, ainsi que ses applications. Les principes de conception d'une technique de tatouage sont ensuite étudiés. Les attaques auxquelles une image tatouée est potentiellement soumise sont classées. On introduit ensuite l'analogie classique avec les télécommunications, qui conduit au tatouage par étalement de spectre, étudié plus en détail dans le chapitre 2. C'est l'analogie avec les problèmes de codage source et de codage canal qui conduit à construire des codes de tatouage, qui constituent la deuxième grande famille de méthodes de tatouage. Des méthodes de tatouage atypiques sont ensuite présentées. Deux techniques de tatouage informé sont plus détaillées : l'étalement de spectre amélioré et les techniques quantificatives, qui seront utiles dans le chapitre 2. Un état de l'art des avancées récentes sur la sécurité des algorithmes de tatouage est ensuite proposé. Enfin, les modèles statistiques et perceptuels liés au tatouage d'images numériques "naturelles" sont présentés.

1.1 Principe du tatouage

L'objet de cette thèse est de développer des techniques de tatouage numérique, dont l'une des principales applications est la protection des droits d'auteurs sur des œuvres numériques. La prolifération des document numériques (notamment par le biais d'internet et des réseaux d'échange de pair à pair ou *P2P*) conduit en effet à une remise en cause de la gestion classique des droits d'auteur. Un utilisateur mal intentionné peut instantanément obtenir une copie d'un document numérique qui est strictement identique à l'original, et en contester la propriété. Le tatouage numérique vise notamment à résoudre de tels conflits.

1.1.1 Historique du droit d'auteur

Principe du droit d'auteur

Deux systèmes de protection des œuvres existent dans le monde : le droit d'auteur (notamment utilisé en France) et le *copyright* (notamment utilisé aux Etats-Unis). La principale différence entre ces deux systèmes réside dans les conditions de protection. En France, elle est implicite dès la création (pas de dépôt formel), aux USA la création doit être tangible (durable) et pour permettre une action en justice, elle doit être déposée au *Copyright Office*. En France, le droit d'auteur est né à la suite de la Révolution en 1791, autour du concept de personnalité unique de l'auteur d'une œuvre [PTB+05]. L'auteur y acquiert un droit de représentation et de reproduction sur son œuvre. On différencie le droit moral (respect de l'intégrité de l'œuvre, droit de retrait), et les droits patrimoniaux (reproduction, distribution), qui perdurent 70 ans après le décès de l'auteur. Les exceptions à la protection sont la liberté d'information (analyses courtes de l'œuvre, reproduction dans une revue de presse) et la liberté de création (parodie). La législation française actuelle s'appuie sur le Code de la Propriété Intellectuelle qui comprend la propriété littéraire et artistique (droits d'auteur) et la propriété industrielle (brevets). Les critères de protection d'une œuvre sont sa concrétisation intellectuelle (d'une idée non protégeable, à une œuvre) et matérielle, mais surtout son originalité.

La spécificité du système des droits d'auteurs est illustrée par la bataille dont font l'objet les logiciels informatiques au Parlement Européen en 2006. Le logiciel est en effet protégé en France par le droit d'auteur depuis 1985, bien que ces droits reviennent

automatiquement à l'employeur. Aux USA, les logiciels font l'objet de brevets, système que la Commission Européenne propose d'imposer en Europe. En effet, un brevet doit être explicitement déposé, éventuellement tenu secret, et est payant pendant toute la durée de la protection [CL04]. L'ampleur de la controverse souligne la différence fondamentale entre ces deux systèmes de protection. Par son caractère implicite, le système du droit d'auteur empêche le dépôt d'une œuvre par quelqu'un d'autre que son auteur. D'autre part, le système du droit d'auteur permet l'existence du "logiciel libre" : l'auteur peut signer une licence (par exemple sur le modèle des licences GPL, pour *General Public License*) dans laquelle il exprime le souhait de ne pas être protégé dans l'exploitation de son œuvre, tout en conservant son droit moral. Sur ce modèle, commence à se développer le courant dit de l'"art libre", avec les licences *Creative Commons* ou les Licences Art Libre.

Cas particulier des documents numériques

Les œuvres numériques posent cependant un problème d'application du droit d'auteur. Une œuvre numérique peut être distribuée de manière légale sous forme concrète (CD, DVD), ou *via* des plates-formes payantes de téléchargement qui permettent la rémunération de l'auteur. Cependant, il est très aisé de fabriquer une copie absolument identique à l'œuvre numérique originale, ainsi que de la distribuer. La contrefaçon ne nécessite pas de moyen technique particulier. Le problème est devenu particulièrement aigu avec l'apparition du système *P2P*). Il s'agit d'un système d'échange de fichiers d'ordinateur à ordinateur qui réunit près de 10 millions d'utilisateurs dans le monde. Si la technologie elle-même n'est pas illicite, le fait de partager des fichiers protégés par le droit d'auteur l'est, puisque les ayant-droits ne sont pas rémunérés lors de l'échange. Le téléchargement est légal, mais la mise en ligne (*upload*) est soumise à autorisation, or dans le *P2P* chaque utilisateur est à la fois émetteur et récepteur. Le droit français ajoute une complication supplémentaire en autorisant la copie privée. Son détournement à des fins de piratage a conduit à la création d'une taxe sur les supports CD et DVD vierges, destinée aux auteurs. En 2006, une nouvelle loi sur les "droits d'auteur dans la société de l'information" a été votée suite à une directive de 2001 du Parlement Européen, et rénove en profondeur la question des droits d'auteurs numériques.

Systèmes de protection des œuvres numériques (DRM)

La loi DADVSI (Droit d'Auteur et Droits Voisins dans la Société de l'Information) a été promulguée en août 2006 [Web06], malgré l'opposition de nombreux partisans d'une "license globale" sur les œuvres copiées. Elle réaffirme le principe de l'exception pour copie privée, mais le téléchargement illégal est désormais passible de 38 euros en cas de flagrant délit et la mise à disposition de 150 euros. Les éditeurs de logiciels d'échanges risquent trois ans d'emprisonnement et 300.000 euros d'amende.

La loi introduit une reconnaissance des systèmes de protection et de contrôle des œuvres numériques (DRM, *Digital Rights Management*, aussi appelés "verrous numériques") en précisant que leur rôle est "d'empêcher ou de limiter les utilisations non autorisées". Les systèmes de DRM actuels sont fondés sur le cryptage du contenu multimédia diffusé, son décryptage nécessitant une licence (éventuellement acquise en ligne). Si le tatouage numérique est une alternative aux systèmes de cryptage, il n'est souvent qu'un chaînon dans un système utilisant cryptographie et communications client-serveur. La protection des ayant-droits n'est pas sans contrepartie sur la liberté de l'usager : on peut craindre notamment que les éditeurs associent DRM et

lecteur multimédia pour imiter le succès d'iTunes d'Apple, qui est à la fois un logiciel propriétaire de gestion de bibliothèque musicale et un lecteur de musique numérique. Les pessimistes imaginent également une possible surveillance des échanges de données par des fichiers "piégés". L'interopérabilité des DRM n'est pas explicitement exigée par la loi DADVSI : les éditeurs de DRM pourront conserver leur code source secret, ce qui ouvre une brèche vers une brevetabilité des programmes de protection. Le contournement des DRM est sanctionné. Le fournisseur de moyens de contournement est passible de 30 000 euros d'amende et de six mois de prison, un particulier ayant décrypté une mesure de protection, de 3750 euros, et un utilisateur de logiciel permettant de contourner les DRM, de 750 euros. L'application stricte de cette loi fait craindre à certains une limitation de la liberté d'expression des journalistes et des chercheurs. Publier une étude démontrant qu'un système de tatouage numérique utilisé par l'industrie du disque est inefficace pourrait faire risquer une peine de prison ferme.

Cette introduction met en avant deux applications essentielles du tatouage numérique : la protection de copie (limiter le nombre de copies privées, interdire la copie à usage d'un tiers) et la gestion des droits d'auteurs. Dans ce dernier cas, rappelons que le tatouage (ou du moins le document original) doit toujours être enregistré auprès d'un tiers de confiance. La reconnaissance des systèmes de protection par la loi DADVSI ouvre la voie à leur généralisation, et le tatouage numérique est un bon candidat. Notons qu'il aurait également pu être utilisé dans le cadre d'une license globale (hypothèse abandonnée pour l'instant), afin de recenser l'utilisation d'une œuvre pour permettre la rémunération de son auteur. Dans un tel cadre, le tatouage aurait été moins sujet aux attaques hostiles.

1.1.2 Présentation du tatouage numérique

Le tatouage numérique, *digital watermarking* en anglais, consiste à insérer un tatouage dans un document numérique (image, son, vidéo…). La modification s'effectue dans les composantes perceptibles (comme la luminance des pixels d'une image), et non dans l'en-tête d'un fichier par exemple. Ce tatouage doit pouvoir être détecté et décodé, mais doit être **imperceptible**, c'est-à-dire que la déformation doit être suffisamment faible pour que l'utilisateur ne puisse pas différencier le document tatoué de l'original. Cette notion d'imperceptibilité et d'insertion dans la trame même du document rejoint la traduction littérale du terme *digital watermark* : "filigrane électronique". On peut trouver les premiers filigranes sur des papiers du treizième siècle, dans le but de garantir leur qualité. Sur un billet de banque, les fibres sont marquées au moment de la sortie du bain d'eau, ce qui est à l'origine du terme anglais *water mark*. De la même manière que sur un billet de banque, le filigrane électronique est d'abord invisible et n'est révélé que par une transformation spécifique. L'intérêt d'une telle opération est que le tatouage est indépendant du format de stockage des données, puisqu'il est intrinsèque au document. C'est donc une solution élégante au vieux problème du "trou analogique" des systèmes de DRM : comment conserver un DRM si l'utilisateur numérise le rendu analogique du document, par exemple par impression/numérisation ? Ainsi, une musique tatouée sur un CD pourra être identifiée même après extraction et compression en mp3. On parle alors de "sécurité au niveau du contenu". Un autre avantage du tatouage sur les systèmes de DRM classiques est qu'il n'est pas obligatoire de recourir à un "tiers de confiance" délivrant les licenses : le document tatoué est lisible par tous les utilisateurs. Le tatouage est une technique de dissimulation d'information (*information hiding*), principe qui englobe également la transmission d'une information secrète dans un réseau ou encore la stéganographie.

Le document tatoué est destiné à être distribué à grande échelle, il est donc amené à subir des déformations. Celles-ci peuvent être involontaires (par exemple : compression d'une image au format JPEG, puis décompression) ou volontaires (pirate voulant endommager le tatouage). La **robustesse** à de telles attaques est l'une des propriétés importantes d'une méthode de tatouage. Les attaques les plus simples (légère rotation ou translation d'une image, rognage de quelques lignes ou colonnes) obtiennent déjà des résultats dévastateurs sur les méthodes initialement imaginées [VPP+01], et les chercheurs ont mis en évidence des attaques beaucoup plus perfectionnées [KP03b]. Le tatouage est même modélisé comme un jeu entre le tatoueur et l'attaquant [MO03].

La troisième contrainte importante du tatouage est la quantité d'information que l'on peut insérer, ou **capacité** : pour une fiabilité de détection donnée, plus l'on insère d'information, plus la déformation est importante. On doit donc trouver un compromis entre trois objectifs antagonistes : imperceptibilité, robustesse et capacité.

La **sécurité**, au sens cryptographique du terme, de la méthode de tatouage, constitue une quatrième contrainte indépendante des trois premières. Elle concerne par exemple la génération de la clé secrète, ainsi que le protocole d'échange général. La méthode de tatouage doit également respecter le principe suivant énoncé par Kerckhoff : l'algorithme lui-même doit pouvoir être rendu public, la sécurité ne dépendant pas de son caractère secret.

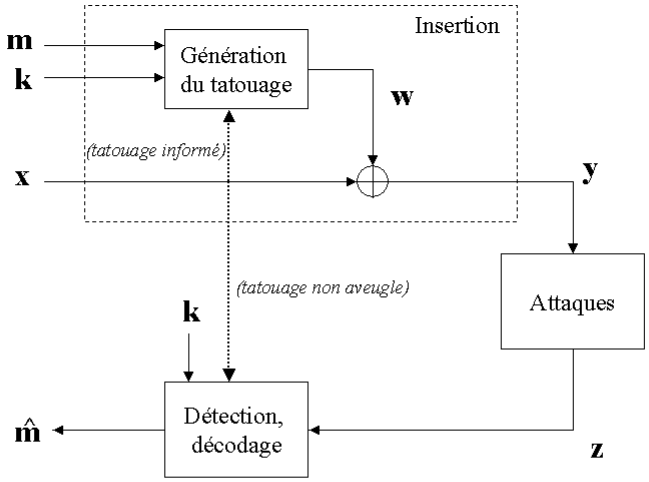

FIG. 1.1 – Principe du tatouage

Le schéma du tatouage numérique est résumé Fig. 1.1 : un message $\mathbf{m}$ contenant L bits d'information est transformé selon une clé $\mathbf{k}$ en un tatouage $\mathbf{w}$ qui est ensuite inséré dans le document $\mathbf{x}$ (aussi appelé "hôte") pour donner un document tatoué $\mathbf{y}$. C'est la phase d'**insertion**. Ici, $\mathbf{w}$ est exprimé sous la forme d'un bruit qui est ajouté au document, la déformation dépendant de la puissance du bruit. $\mathbf{k}$ est secrète et spécifique au tatoueur. $\mathbf{y}$ est ensuite copié et attaqué, ce qui est modélisé par la transmission dans

un canal soumis à du bruit. Le document reçu est appelé $\mathbf{z}$. La réception d'un document consiste en deux parties : d'une part, la détection du tatouage et d'autre part, s'il est présent, son décodage. La phase de **détection** consiste à prouver la présence d'un tatouage dans $\mathbf{z}$ grâce à $\mathbf{k}$. La phase de **décodage** consiste à calculer une estimation $\hat{\mathbf{m}}$ de $\mathbf{m}$. Si la taille du message inséré L est suffisamment grande et contient une information intelligible (par exemple, des caractères ASCII), certains auteurs considèrent que la détection devient inutile puisqu'on peut appliquer un simple décodage. Si la chaîne décodée est inintelligible (par exemple, non ASCII), on considère qu'il n'y a pas de tatouage [MDC04][CMB02].

Si le document original n'est pas utilisé à la réception, l'algorithme de tatouage est qualifié d'**aveugle**. Dans le cas inverse (beaucoup moins intéressant en pratique), l'algorithme est qualifié de non aveugle ou à décodeur informé. Si le document original est utilisé dans la construction de $\mathbf{w}$, on parlera de **tatouage informé**. Lorsque plusieurs tatouages sont insérés (correspondant souvent à plusieurs utilisateurs), on parle de **tatouage multiple**.

1.1.3 Applications

Si le terme *digital watermarking* a été introduit en 1990, l'explosion du nombre de publications à ce sujet date de 1995, ce qui s'est concrétisé par la création de l'atelier IHW (*Information Hiding Workshop*) en 1996, d'une conférence spécifique au sein de SPIE en 1999 et de l'atelier IWDW (*International Workshop on Digital Watermarking*) en 2002. Quatre journaux dédiés aux problématiques de sécurité de l'information ont été crées récemment : *IEEE Trans. on Information Forensics and Security* et *IEE Proc. Information Security* en 2005, *LNCS Transactions on Data Hiding and Multimedia Security* et *EURASIP Journal on Information Security* en 2006, ce qui souligne le dynamisme du domaine. Les projets français (AQUAMARS : 1999-2001, AQUAFLUX, sur le tatouage de flux multimédia : 2002-2003, ARTUS : 2002-2005), européens (Certimark, sur l'évaluation d'algorithmes de tatouage : 2000-2002) et le réseau d'excellence européen ECRYPT (dont le laboratoire virtuel WAVILA) fédèrent certaines recherches.

Les promesses du tatouage ont conduit à la prolifération d'entreprises dans le domaine, même si l'enthousiasme initial semble retombé. Digimarc, firme pionnière, rassemble des brevets de base sur le tatouage (notamment celui de l'estampillage, défini plus loin) dont elle vend la licence. Elle est également auteur du module de tatouage du logiciel de traitement d'image Photoshop. Son concurrent Verance fournit les outils de contrôle de flux audiovisuel Broadcast Verification et ConfirmMedia. La compagnie Liquid Audio fournit également un système de tatouage audio. Le SDMI (Secure Digital Music Initiative) est un consortium de compagnies pour un projet de tatouage audio. Les associations japonaises JASRAC et RIAS sont également actives dans ce domaine. En France, Nextamp et MediaSec, filiales de Thomson, s'intéressent au suivi et à la sécurité vidéo. Notamment, l'Institut National de l'Audiovisuel (INA) utilise le système de tatouage vidéo de Thomson pour une application de suivi des transactions : le document téléchargé contient le nom de l'acheteur. L'institut Fraunhofer (créateurs du mp3) a annoncé en 2006 avoir développé un logiciel de tatouage audio commercialisable.

Le tatouage numérique a donc de nombreuses applications, dont l'une des plus porteuses est la **gestion des droits d'auteur numériques**. Certaines législations imposent de déposer les droits d'un document auprès d'un tiers de confiance, qui délivre ensuite un identifiant. C'est cet identifiant qui sera tatoué dans le document. Dans le domaine de la **protection de copie**, on détecte la présence d'un *copyright* sur un document, dans le but d'empêcher sa manipulation par exemple. Cet enjeu est très important

pour les maisons de disques regroupées dans le consortium SDMI face à l'échange de mp3. En pratique, on peut par exemple imaginer un environnement logiciel et matériel totalement compatible avec le tatouage, qui empêche la copie d'un document tatoué. Une autre application est la **gestion des transactions** (*transaction tracking*) ou estampillage (*fingerprinting*) pour laquelle on insère l'identité du vendeur et celle de l'acheteur. Les propriétaires successifs du document, et donc les sources de copie d'un document peuvent ainsi être identifiés. Un schéma de tatouage multiple est nécessaire.

Ces méthodes peuvent être regroupées sous l'appellation de **tatouage robuste**, car ils doivent être opérationnels même en cas d'intervention d'attaques malveillantes. La puissance et la diversité des attaques répertoriées font qu'aucun système de tatouage totalement viable n'a pour l'instant été adopté. Les normes s'orientent vers des méthodes de tatouage simples, nécessitant peu de calculs, même si elles ne sont pas sûres ou peu robustes, dans le but de décourager une majorité d'utilisateurs, à l'exemple de la proposition du consortium Millenium pour la protection des DVD.

D'autres champs d'intérêt englobent le **tatouage sans perte** [SNZ$^+$04][FG02], ou tatouage réversible [FG01], où l'on désire pouvoir récupérer de façon exacte le document initial, et le **tatouage fragile**. Dans ce dernier cas, le tatouage est volontairement vulnérable aux attaques dans le but de détecter une manipulation éventuelle du document. On peut ainsi justifier auprès d'un tribunal l'authenticité de documents tels que des enregistrements de caméra de surveillance (**authentification de contenu**). L'intérêt d'une technique de tatouage fragile dépend entre autres de la possibilité de localiser les zones de l'image manipulées, ou encore la manipulation effectuée. On parle alors de tatouage révélateur (*tell-tale watermarking*).

Le **tatouage légiste** (*forensic watermarking*) regroupe les applications qui peuvent directement entraîner l'intervention des tribunaux : authentification, tatouage fragile pour authentification d'un témoignage ou de la validité d'un chèque. Il inclut notamment un scénario proche de l'estampillage, dans lequel on autorise les copies d'un document, tout en pouvant remonter à la source du piratage. Le but est alors d'attaquer le pirate en justice, à titre dissuasif pour les autres utilisateurs. Ce scénario est souvent évoqué pour un contenu musical ou pour le cinéma en ligne. Le **tatouage semi-fragile** [LLH05] vise à résister à certains traitements du document, tant que son contenu sémantique n'est pas altéré. On distingue les attaques légitimes (ex : compression JPEG), auxquelles la méthode est robuste, des attaques illégitimes auxquelles elle est fragile.

On peut également transmettre secrètement un message. Ce domaine d'application a des contraintes légèrement différentes et s'appelle la **stéganographie**. Dans ce cas de figure, le document est transmis d'utilisateur à utilisateur et moins exposé à des attaques. Les contraintes de sécurité, d'imperceptibilité et de capacité prennent donc le pas sur celle de robustesse, et beaucoup de techniques de stéganographie sont basées sur la modification des composantes les moins perceptibles du document.

Enfin, un tatouage peut servir à insérer une information supplémentaire dans le document, sans contrainte de sécurité ou de robustesse : il s'agit d'amélioration de contenu, qui est donc l'application la plus viable. On peut ainsi ajouter des informations sur l'artiste dans une chanson diffusée à la radio, ou dans l'application de "contrôle d'appareil", être dirigé vers un site internet en scannant une publicité magazine (DigiMarc Media-Bridge). On parle de **documents auto-indexés** lorsque le tatouage contient sa propre description, afin de permettre son stockage dans une base de données sans problème de changement de format. Les **contenus augmentés** peuvent également servir à ajouter le nom de l'interprète d'une chanson ou une traduction en langage des signes dans un document vidéo.

Dans le cadre du tatouage multiple de J tatouages, il existe deux façons de trans-

mettre L_0 bits d'information. La première est d'insérer L_0 tatouages de 1 bit ($J = L_0$, $L = 1$) orthogonaux entre eux. On peut également insérer 1 message de L_0 bits ($L = L_0$, $J = 1$). Le tatouage multiple permet également d'utiliser d'autres valeurs de J et L avec $L_0 = JL$. Le tatouage multiple est effectué soit de manière simultanée et indépendante, soit de manière séquentielle en utilisant l'information disponible sur les clés secrètes utilisées précédemment [WCA04]. L'intérêt du tatouage multiple est large : contrôle des copies d'un document, suivi des transactions... De plus, chaque application du tatouage a ses propres contraintes (ex : tatouage fragile/robuste), ce qui peut amener à insérer plusieurs tatouages pour différents usages.

La quantité d'information insérée dans le tatouage, ou "charge utile" (*payload*), est très variable selon les algorithmes et applications proposés. Pour l'insertion d'un copyright, on peut par exemple vouloir insérer une information similaire à la norme ISBN utilisée pour les livres, soit de 60 à 70 bits d'information [LSL00]. Un message encore plus long peut être inséré si l'image est considérée comme un canal de communication caché. Une charge utile de $L = 1000$ ou plus est parfois envisagée [MDC04]. A l'inverse, beaucoup d'auteurs proposent d'insérer une information binaire (présence du tatouage ou non), pour une application à la protection de copie. On parle également de **signature** pour la séquence ainsi générée. Dans la suite, on étudiera principalement le cas de la communication de messages. La vérification de signature sera donc considérée comme un cas particulier où l'on se limite à la détection. Si les performances de détection requises varient énormément selon l'application, on trouve néanmoins dans [CMB02] les exemples des normes DVD : $P_{\text{fa}} = 10^{-6}$ pour la preuve de propriété et $P_{\text{fa}} = 10^{-12}$ pour la protection de copie.

Les travaux de cette thèse ne sont pas dirigés vers une application particulière. Les techniques proposées sont conçues comme des alternatives ou des variantes des techniques générales de tatouage existantes. Cependant, on s'intéressera plus particulièrement au tatouage robuste (de préférence au tatouage fragile, à la stéganographie ou au tatouage sans perte par exemple, qui impliqueraient d'autres contraintes). Le tatouage multiple sera évoqué, ce qui permet une application à l'estampillage. Les techniques proposées sont applicables à tout signal numérique. A l'exception de quelques modèles statistiques, on ne prend pas en compte la structure spécifique d'un type de document, contrairement aux algorithmes de tatouage d'image 3D, *halftone* ou encore de code informatique. Toutefois, les applications pratiques que nous considérons sont réalisées sur des images naturelles.

1.1.4 Perspectives pour le tatouage numérique

Les attentes suscitées par le tatouage jusqu'à une date récente ont été très élevées (notamment lorsqu'il était considéré comme un remède miracle contre le piratage) et pour l'instant déçues. Une controverse a été lancée sur l'utilité du tatouage, notamment par C. Herley [Her02]. Il y affirmait que la variété des attaques envisageables sur un document tatoué est bien plus grande que les cas de figure traités jusqu'ici. Ainsi, beaucoup d'images ne perdent pas leur signification après des rotations de 90, 180 ou 270 degrés, ou des modifications de couleur de certains objets. D'après Herley, l'ensemble des images attaquées n'est donc pas connexe et on ne peut limiter les déformations acceptables à un voisinage de l'image, comme c'est le cas en tatouage. De plus, l'échec de l'appel lancé par le SDMI (consortium de l'industrie du disque) afin de créer un système de tatouage audio sûr et robuste a conduit les chercheurs à abandonner l'objectif de sécurité totale face à des attaques. Aucun des algorithmes actuels n'est par exemple robuste à toutes les attaques géométriques locales non affines.

Cette controverse a poussé la communauté de chercheurs à apporter des réponses sur les applications viables du tatouage et à faire un bilan des techniques les plus prometteuses [Bar03][Mou03]. Les applications les moins orientées sur la sécurité sont les plus exploitables à l'heure actuelle : contrôle de diffusion (d'une chanson par une radio, par exemple), tatouage fragile, amélioration de contenu (où l'on peut exploiter la propriété de survie du tatouage à des changements de format). M. Barni suggère également de généraliser le principe de la taxe appliquée en France sur les CDs vierges et destinée aux maisons de disques [BB04]. Chaque support de stockage serait taxé à l'achat et le tatouage servirait à contrôler le nombre de copies effectuées d'un document, dans le but d'attribuer des droits d'auteur en proportion. On déplacerait donc le problème du piratage de l'utilisateur vers le propriétaire, supposé plus facile à surveiller.

Cependant, le tatouage numérique est un domaine jeune qui progresse rapidement. Si la robustesse aux attaques géométriques reste le talon d'Achille des techniques de tatouage, de nouvelles solutions continuent d'être proposées [DBG$^+$05][DRRD06]. Parmi les perspectives prometteuses, on trouve également l'utilisation de techniques récentes de codage (récemment formalisée par P. Moulin [MK05]) et leurs liens avec le tatouage quantificatif à haute dimension, ou encore l'optimisation sous contraintes. De grands progrès ont été faits ces dernières années sur la sécurité des algorithmes de tatouage [CFF05]. La cryptographie jointe au tatouage et la compression jointe au tatouage ont également suscité un intérêt récent. Par exemple, au lieu d'utiliser une clé secrète pour la sécurité, puis de tatouer, puis de compresser, il est préférable d'effectuer un codage conjoint pour le tatouage, la compression avec perte et le cryptage [Mer05].

1.2 Conception d'une méthode de tatouage

Cette partie présente un état de l'art des principales techniques de tatouage, sous l'angle de leur réponse aux contraintes du problème et de leur inspiration multidisciplinaire. Dans un premier temps, des notations générales sont introduites. L'ensemble des attaques à prendre en compte est répertorié. Les trois principales stratégies d'insertion sont ensuite présentées. Les techniques de tatouage additif par étalement de spectre, le principe du tatouage informé et les techniques de tatouage substitutif sont développés. Nous nous intéressons ensuite à l'apport d'autres disciplines. Dans la littérature, on pourra se reporter aux états de l'art de F. Hartung et M. Kutter [HK99] ainsi qu'au livre [DP04], et plus récemment à l'approche théorique de P. Moulin [MK05].

1.2.1 Notations

$\mathbf{x}$ désigne le document original, $\mathbf{w}$ le tatouage, $\mathbf{y}$ le document tatoué, $\mathbf{k}$ la clé, $\mathbf{m}$ le message. Pour le tatouage multiple à J utilisateurs, on associe $\mathbf{w}^j$ et $\mathbf{m}^j$ à l'utilisateur j. Pour plus de simplicité, on utilisera des messages de taille L binaires antipodaux ($m_l = \pm 1$). Les documents sont considérés soit comme des matrices de taille $N_1 \times N_2$ (cas d'une image) :

$$\begin{aligned} \mathbf{w}^j &= [w^j_{k_1,k_2}]_{k_1\in\{1,\dots,N_1\},k_2\in\{1,\dots,N_2\}} \\ \mathbf{x} &= [i_{k_1,k_2}]_{k_1\in\{1,\dots,N_1\},k_2\in\{1,\dots,N_2\}} \\ \mathbf{y} &= [y_{k_1,k_2}]_{k_1\in\{1,\dots,N_1\},k_2\in\{1,\dots,N_2\}} \ , \end{aligned} \tag{1.1}$$

soit comme des vecteurs de taille $N = N_1 N_2$ comme suit :

$$\begin{aligned} \mathbf{m}^j &= [m_l^j]_{l\in\{1,\dots,L\}} \\ \mathbf{w}^j &= [w_k^j]_{k\in\{1,\dots,N\}} \\ \mathbf{x} &= [x_k]_{k\in\{1,\dots,N\}} \\ \mathbf{y} &= [y_k]_{k\in\{1,\dots,N\}} \end{aligned} \tag{1.2}$$

Pour une image, on passe de la notation matricielle à la notation vectorielle par défaut en concaténant les lignes (les pixels sont donc pris en ordre "lexicographique") ou en utilisant un parcours d'image (dans ce cas le parcours est précisé, cf. paragraphe 2.2).

Pour l'utilisateur, $\mathbf{w}$ est un bruit de faible amplitude. Pour l'encodeur, $\mathbf{w}$ est le signal intéressant. L'algorithme de tatouage équivaut donc à la transmission de $\mathbf{w}$ à travers un canal très bruité. Certaines attaques (cf. partie 1.2.3) sont modélisées par une source de bruit simple $\mathbf{n} = [n_k]_{k\in\{1,\dots,N\}}$. On définit les rapports signal à bruit d'insertion DWR (*Document to Watermark Ratio*) et de transmission WNR (*Watermark to Noise Ratio*), ainsi que le rapport document à bruit DNR (*Document to Noise Ratio*). Dans le cadre aléatoire, le document, le tatouage et le bruit sont des variables aléatoires X, W et N. En notant $\sigma^2_{\mathbf{x}}$, $\sigma^2_{\mathbf{w}}$ et $\sigma^2_{\mathbf{n}}$ les variances respectives de chacun de leurs éléments,

$$\text{DWR} \triangleq \frac{\sigma^2_{\mathbf{x}}}{\sigma^2_{\mathbf{w}}},\ \text{WNR} \triangleq \frac{\sigma^2_{\mathbf{w}}}{\sigma^2_{\mathbf{n}}},\ \text{DNR} \triangleq \frac{\sigma^2_{\mathbf{x}}}{\sigma^2_{\mathbf{n}}}\ .$$

$\mathbf{w}$ et $\mathbf{n}$ seront toujours de moyenne nulle. On estime donc en pratique ces rapports par :

$$\text{DWR} = \frac{\sum_{k=1}^N (x_k - \mu(\mathbf{x}))^2}{\sum_{k=1}^N w_k^2},\ \text{WNR} = \frac{\sum_{k=1}^N w_k^2}{\sum_{k=1}^N n_k^2},\ \text{DNR} = \frac{\sum_{k=1}^N (x_k - \mu(\mathbf{x}))^2}{N\sigma^2_{\mathbf{n}}}\ .$$

La performance au décodage est mesurée par le taux d'erreur bit (TEB) défini par :

$$\text{TEB} \triangleq p[d_l \neq m_l] = \frac{1}{2}\left(\text{p}[\hat{m}_l = 1 | m_l = -1] + \text{p}[\hat{m}_l = -1 | m_l = 1]\right)\ ,$$

où $\hat{\mathbf{m}}$ est la décision finale sur l'estimation du message et en supposant que les probabilités *a priori* pour que $m_l = 1$ et $m_l^j = -1$ sont égales. Lors des simulations et dans le cadre du tatouage multiple, le TEB est estimé par :

$$\widehat{\text{TEB}} = \frac{\sum_{j=1}^J (1 - \sum_{l=1}^L \delta(\hat{m}_l^j, m_l^j))}{JL}\ ,$$

où δ désigne le symbole de Kronecker. Une autre manière d'évaluer un algorithme de tatouage lorsqu'on dispose de beaucoup de documents est de calculer le pourcentage de documents dans lesquels le tatouage a été parfaitement décodé [MDC04].

1.2.2 Mise en forme du message

On appelle "mise en forme" l'opération qui fait passer du message $\mathbf{m}$ de taille L au message redondant $\mathbf{b}$ de taille N, et vérifiant : $b_n = m_l \quad \forall n \in \mathcal{S}_l$. $\mathcal{S}_l$ est l'ensemble de cardinal P des coordonnées correspondant au bit l, et $\cup_{i=1}^L \mathcal{S}_i = \{1,\dots,N\}$. Dans la suite, pour un vecteur $\mathbf{x}$ donné de longueur $N = LP$, $\underline{\mathbf{x}}_{\mathbf{l}}$ signifiera $[x_k]_{k\in\mathcal{S}_l}$, avec $l \in \{1,\dots,L\}$. $P \triangleq N/L$ désigne donc la redondance de chaque bit d'information. Le débit du message est défini par $R \triangleq L/N = 1/P$ en bit/échantillon.

Dans la **mise en forme NRZ** (Non Retour à Zéro), $\mathcal{S}_l = \{(l-1)P+1, \ldots, lP\}$. Du fait des contraintes matérielles (notamment, la modulation en temps réel du message), c'est la mise en forme la plus utilisée en télécommunications. En tatouage, cette mise en forme peut être utile en cas de tatouage en temps réel (d'un flux audio par exemple). Cependant, dans la plupart des applications, et en particulier en image où le document est de taille finie, le tatoueur manipule directement des signaux de taille N. On utilise alors le plus souvent une **mise en forme aléatoire** [Har99][HPGRN98] : les P échantillons correspondant au bit l constituent un ensemble $\mathcal{S}_l$ de points répartis aléatoirement sur l'image. Cela correspond à appliquer un entrelaceur aléatoire de taille $N \times N$ après une mise en forme NRZ. On introduira la **mise en forme répétition** dans la partie 2.1.2 : elle consiste à utiliser $\mathcal{S}_l = \{l, l+L, \ldots, N-L+l\}$.

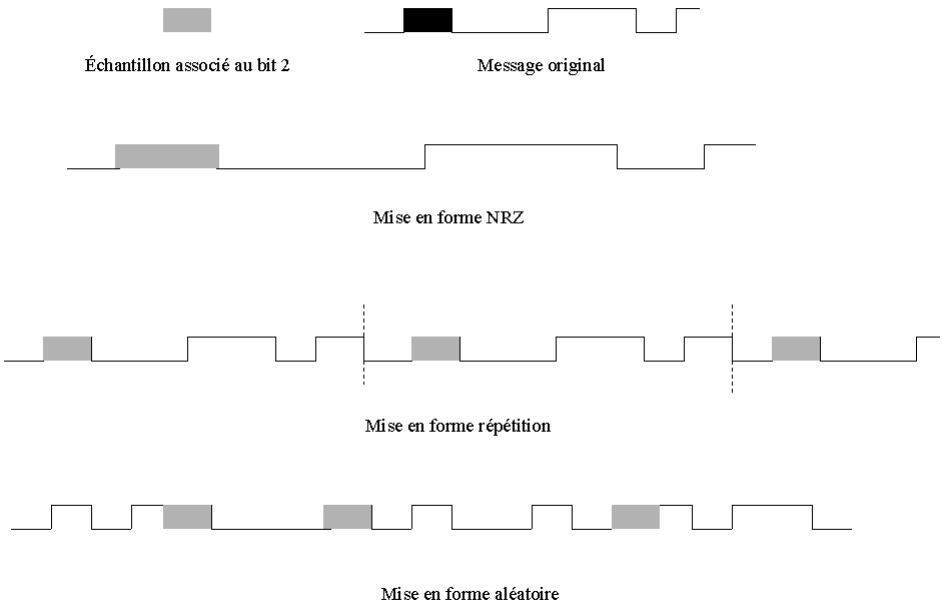

FIG. 1.2 – Différentes mises en forme

1.2.3 Classification des attaques sur la robustesse

Les attaques $\mathcal{A}$ transformant $\mathbf{y}$ en un document attaqué $\mathbf{z}$ sont très variées. On adoptera la classification des attaques de [VPP+01] qui différencie les attaques visant à enlever le tatouage (attaques d'effacement), à déformer suffisamment $\mathbf{y}$ pour rendre la détection impossible (attaques géométriques), à décrypter $\mathbf{k}$ (attaques cryptographiques), et celles visant à trouver une faille dans le protocole de gestion des droits d'auteurs lui-même (attaques de protocole). Les deux premiers types d'attaques peuvent être considérés comme des attaques sur la robustesse, alors que les suivants sont des attaques sur la sécurité, et qui seront développées dans le paragraphe 1.4.1.

Attaques d'effacement

On peut modéliser par une source de bruit simple $\mathbf{n} = [n_k]_{k\in\{1,\dots,N\}}$ les distorsions introduites aussi bien par le canal de transmission que par certaines attaques d'effacement simples (ou *waveform attacks* [Har99]) : insertion d'un bruit additif ou multiplicatif (appelé *speckle*), opération de filtrage. Le modèle classique de canal blanc gaussien (**AWGN**) est généralement utilisé :

$$\mathbf{z} = \mathbf{y} + \mathbf{n} \text{ où : } n_k \sim \mathcal{N}(0, \sigma^2_{\mathbf{n}}) \ . \tag{1.3}$$

Ce modèle est particulièrement adapté aux algorithmes fondés sur une transformée d'étalement, grâce au Théorème Central-Limite. Une extension de ce modèle [PGCB03] s'appuie sur un canal à bruit additif probabiliste $N = \{N^k\}$, où $z_k = y_k + N^k$ avec $N^k \sim \mathcal{N}(0, \sigma^2_{N^k})$, donc $\sigma^2_{\mathbf{n}} = \frac{1}{N}\sum_{n=1}^{N} \sigma^2_{N^k}$.

Lorsque des attaques plus fortes sont effectuées, le modèle de bruit peut être plus compliqué avec par exemple des distributions non gaussiennes. Par conséquent, l'influence de telles attaques sur la performance au décodage est généralement étudiée au travers de simulations. C'est le cas d'attaques d'effacement plus évoluées telles que le **débruitage par filtrage de Wiener**, qui a pour but de séparer le signal $\mathbf{x}$ du "bruit" **w**, le document étant supposé blanc et gaussien [SEG01]. En image, la **compression JPEG** consiste à quantifier les coefficients de la DCT par blocs 8x8 (cf. paragraphe 1.5.1) de $\mathbf{y}$. Le pas de quantification y varie pour chaque coefficient en fonction de son importance perceptuelle et du facteur de qualité. Certains auteurs proposent de compresser **w** avant insertion, en compensant la perte d'énergie due à la compression, afin d'augmenter la robustesse à la compression JPEG à DWR donné dans le domaine spatial. Un modèle théorique de l'impact de la compression JPEG sur les performances au décodage a été calculé dans [EG01][FKK04]. La compression JPEG2000 est une attaque destinée à devenir courante, avec de nombreux paramètres (taux de compression, noyau d'ondelette choisi...) [FS02].

Les **transformations valumétriques**, fréquentes en traitement d'images mais surtout en vidéo et en audio (ex : simples changements de volume), incluent par exemple l'égalisation d'histogramme ou encore la correction gamma :

$$\Gamma_\gamma(\mathbf{y}) = \max(\mathbf{y}) \left(\frac{\mathbf{y}}{\max(\mathbf{y})} \right)^\gamma$$

En particulier, on appelle **attaque de gain** la combinaison d'une multiplication par un scalaire et d'un ajout de bruit gaussien, ce qui peut modéliser certains filtrages linéaires ainsi que l'égalisation d'histogramme.

Attaques géométriques

Les attaques géométriques peuvent empêcher la détection du tatouage : légère rotation ou translation d'image, changement d'échelle... En image, on peut modéliser des attaques géométriques simples (non locales), en posant (t_1, t_2) les coordonnées continues correspondant aux pixels d'origine (k_1, k_2), et $A = \{a_{ij}\}$, τ_h, τ_ν les paramètres de la transformation [ARPG02] :

$$\begin{pmatrix} t_1 \\ t_2 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} n_1 \\ n_2 \end{pmatrix} + \begin{pmatrix} \tau_h \\ \tau_\nu \end{pmatrix}$$

Par exemple, une rotation d'angle θ est modélisée par $A(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$ et $\tau_h = 0, \tau_\nu = 0$. Un changement d'échelle correspond à $A(\rho_x, \rho_y) = \begin{pmatrix} \rho_x & 0 \\ 0 & \rho_y \end{pmatrix}$ et $\tau_h = 0, \tau_\nu = 0$. Pour ces deux transformations, une interpolation est nécessaire. Le rognage de lignes ou de colonnes (*cropping*) et la translation correspondent à $A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ et τ_h, τ_ν variables. On rencontre également le cisaillement (*shearing*), qui consiste à étirer l'image selon l'un de ses axes [DBG+05] : $A = \begin{pmatrix} 1 & a_{12} \\ a_{21} & 1 \end{pmatrix}$. Ces attaques géométriques simples peuvent être combinées en attaques géométriques locales, comme l'attaque Stirmark [Sti]. Celle-ci peut être modélisée par un maillage déformable [DBG+05] (cf. paragraphe 1.5.4). Les attaques géométriques agissent de deux manières : avec des paramètres suffisamment grands, elles introduisent une très grande déformation lorsqu'on compare point par point avec l'image originale (avec le DNR par exemple). La faisabilité de l'attaque repose alors sur l'hypothèse que le contenu de l'image n'est pas affecté et que l'utilisateur ne sera pas gêné. D'autre part, il est possible de conserver un DNR élevé en utilisant des attaques très faibles. L'effet sur le tatouage sera principalement dû à la phase d'interpolation (cf. paragraphe 3.2).

FIG. 1.3 – Exemple d'attaque par cisaillement horizontal : Lena, $t_{12} = 0.05$, $t_{21} = 0$

Autres classifications

Du point de vue de la théorie de l'information, une autre classification est proposée par P. Moulin [MK05], modélisant les attaques par un canal de communication. Ce canal est sans mémoire si l'attaque est indépendante pour chaque pixel (exemple : AWGN), sans mémoire par blocs (exemple : compression JPEG), à régularité statistique (bruit stationnaire, filtrage invariant…). Les attaques les plus difficiles à modéliser (géométriques…) sont qualifiées d'arbitraires.

Devant la diversité des algorithmes et des attaques proposés, un outil générique de test de robustesse a été proposé par F. Petitcolas avec le logiciel StirMark [Pet00][Sti] implanté en langage C++. Un équivalent existe aussi en Matlab avec l'outil Checkmark développé par S. Peireira *et al.* au sein du projet européen Certimark [PVM+01][Che]. Checkmark a pour but d'aider au développement d'algorithmes, tandis que StirMark est

destiné à terme à devenir un outil de certification des performances. Plus récemment, un outil d'évaluation libre de droits appelé OpenWatermark, utilisant la technologie Java, a été proposé [MM06a].

En outre, Petitcolas [Pet00] donne un exemple de profil d'évaluation d'un algorithme de tatouage, avec les paramètres d'attaques conseillés (extrait) :

Niveau de l'attaque	Zéro	Faible	Modéré
Compression JPEG : facteur de qualité	100-90	90-75	75-50
Correction Gamma		0.7-1.2	0.5-1.5
Changement d'échelle		1/2-3/2	1/3-2
Rotation		$\pm$ 0-2	$\pm$ 0-5, 90
Filtre moyenneur			3x3

Les attaques prises en compte sont un élément très important dans la conception d'un algorithme de tatouage. Souvent, la robustesse à l'attaque AWGN sert lors de la conception de l'algorithme, car elle permet de modéliser un grand nombre d'attaques et est le modèle le plus approprié lorsque l'attaque est inconnue. Cependant, d'après ce modèle, certains algorithmes sont inutiles (par exemple, DS en comparaison avec QIM, voir plus loin), alors que leurs propriétés de robustesse sont en réalité précieuses. *A contrario*, si l'on se limite à une attaque particulière (ex : robustesse aux attaques géométriques), le système risque d'être peu performant face aux autres attaques.

1.2.4 Principes d'insertion

La phase d'insertion désigne l'opération qui consiste à passer d'une image $\mathbf{x}$ et d'un tatouage $\mathbf{w}$ à une image tatouée $\mathbf{y}$. On identifie ici trois principaux types d'insertion : l'insertion additive, substitutive avec dictionnaire et substitutive avec contraintes. Ils diffèrent également par le principe de décodage qui leur est associé.

Dans le **tatouage additif**, l'opération est décrite par

$$\mathbf{y} = \mathbf{x} + \mathbf{w} \tag{1.4}$$

$\mathbf{w}$ est supposé ici avoir été généré à partir de $\mathbf{m}$ et de la clé $\mathbf{k}$ par un opérateur de génération que nous appellerons g : $\mathbf{w} = g(\mathbf{m}, \mathbf{k})$. Le document $\mathbf{x}$ peut cependant intervenir dans la génération de $\mathbf{w}$, soit par l'emploi d'un masque perceptuel, soit dans une adaptation au tatouage informé : alors $\mathbf{w} = g(\mathbf{m}, \mathbf{k}; \mathbf{x})$. On parle parfois d'**insertion multiplicative**. Dans ce cas, le tatouage de moyenne 1 est multiplié au document : $\mathbf{y} = \mathbf{x}g(\mathbf{m}, \mathbf{k})$. Cependant, ce schéma peut être assimilé au schéma additif en prenant $\mathbf{w} = g(\mathbf{m}, \mathbf{k}) - 1$:

$$\mathbf{y} = \mathbf{x} + \mathbf{w}\mathbf{x}$$

L'opération de décodage du tatouage est considérée comme une inversion de l'opération g malgré l'ajout de $\mathbf{x}$: nous la noterons g^{-1} : $\hat{\mathbf{m}} = g^{-1}(\mathbf{z}, \mathbf{k})$. La plus populaire des fonctions g utilisées est la technique d'étalement de spectre.

On appelle tatouage substitutif un algorithme dont le principe est de remplacer un élément du signal original par un signal tatoué, ce qui correspond principalement à deux comportements. Le premier, que nous appellerons **tatouage substitutif avec dictionnaire**, consiste à remplacer le signal original par un mot issu d'un dictionnaire noté $\mathcal{M}$. $\mathcal{M}$ est découpé en sous-dictionnaires $\mathcal{M}_{\mathbf{m}}$, chacun correspondant à un message possible $\mathbf{m}$: $\mathcal{M} = \mathcal{M}_1 \cup \cdots \cup \mathcal{M}_{2^L}$. Afin de respecter la contrainte d'imperceptibilité et de transmettre un message, le mot de code inséré doit dépendre à la fois du message secret et du signal original : $\mathbf{y} = g(\mathbf{m}, \mathbf{k}; \mathbf{x})$. Ici, l'opérateur de génération est un

opérateur de choix :

$$g(\mathbf{m}, \mathbf{k}; \mathbf{x}) \in \mathcal{M}_{\mathbf{m}}$$

Le décodage consiste ici à retrouver le dictionnaire auquel appartient $\mathbf{z}$: $\hat{\mathbf{m}}|\mathbf{z} \in \mathcal{M}_{\hat{\mathbf{m}}}$. La plus populaire des techniques répondant à cette définition est la technique quantificative (cf. paragraphe 1.3.2).

Le second type d'algorithmes substitutifs consiste à imposer un ensemble de contraintes aux données marquées [Gue03] : nous le nommerons **tatouage substitutif avec contraintes**. Dans ce cas, on se fixe un opérateur objectif c à résultat booléen. Le but est d'avoir $c(\mathbf{y}, \mathbf{m}) = 1$. Contrairement au cas précédent, on ne s'appuie pas sur un dictionnaire prédéfini mais la démarche est de faire tendre $\mathbf{x}$ vers l'ensemble des images respectant la contrainte par une opération d'insertion g :

$$\mathbf{y} = g(\mathbf{m}, \mathbf{k}, c; \mathbf{x})$$

A la réception, on vérifie simplement que la contrainte est respectée : $c(\mathbf{z}, \hat{\mathbf{m}}) = 1$. C'est le type d'algorithme qui regroupe le plus de techniques atypiques, notamment par la variété des distorsions g applicables à l'image.

Ces définitions sont bien entendu non-exclusives. Par exemple, on peut considérer qu'un tatouage substitutif avec dictionnaire impose la contrainte d'appartenance à un dictionnaire donné. Au niveau purement algorithmique, toute technique substitutive peut également s'interpréter comme une insertion additive, le tatouage additif étant tout simplement $\mathbf{w} = g(\mathbf{k}, \mathcal{M}_{\mathbf{m}}; \mathbf{x}) - \mathbf{x}$. La distinction proposée recouvre donc surtout le principe général de conception de la technique d'insertion et de décodage.

La *fig.* 1.4 illustre deux stratégies d'insertion. Dans le tatouage additif à insertion et décodage aveugles, on ajoute à l'image un tatouage afin d'atteindre la région de détection correspondant au détecteur choisi. La détection ne dépend que de la puissance du tatouage inséré. Plus le tatouage est puissant (donc la détection performante), plus la distorsion est grande. Avec une technique de tatouage avec dictionnaire, il existe plusieurs régions de détection pour un message **m** donné. Le tatouage consiste à pousser $\mathbf{x}$ vers la région de détection la plus proche perceptuellement.

1.2.5 Tatouage additif par étalement de spectre

Lien avec les télécommunications

A partir du modèle de l'équation (1.4), on peut considérer le tatouage comme la transmission d'un signal (le tatouage) dans un canal bruité (le document). Cette analogie avec les télécommunications est à la base de l'utilisation de la théorie de l'information (notamment dans les calculs de capacité) et des techniques de tatouage par étalement de spectre (cf. paragraphe 1.2.5). Le tatouage devient alors une mise en forme du message, suivie d'une modulation. La principale différence avec les télécommunications réside dans l'inversion du rapport Signal/Bruit : ici, la puissance du bruit ($\mathbf{x}$) est beaucoup plus grande que celle du signal ($\mathbf{w}$, imperceptible). Le nombre d'échantillons par bit d'information transmis, appelé redondance, est donc essentiel : plus la redondance est grande, meilleures sont les performances. Si le tatoueur utilise la connaissance de $\mathbf{x}$ à l'insertion (tatouage informé), il s'agit d'une transmission avec information de bord. Pour les télécommunications multi-utilisateurs, les principaux problèmes viennent des Interférences Multi-Utilisateurs (MAI), des interférences multi-trajets (réception simultanée de versions décalées dans le temps d'une même communication), du bruit introduit par le canal et des désynchronisations. Il est impossible pour un utilisateur mal intentionné de substituer un autre signal à une communication. Dans le

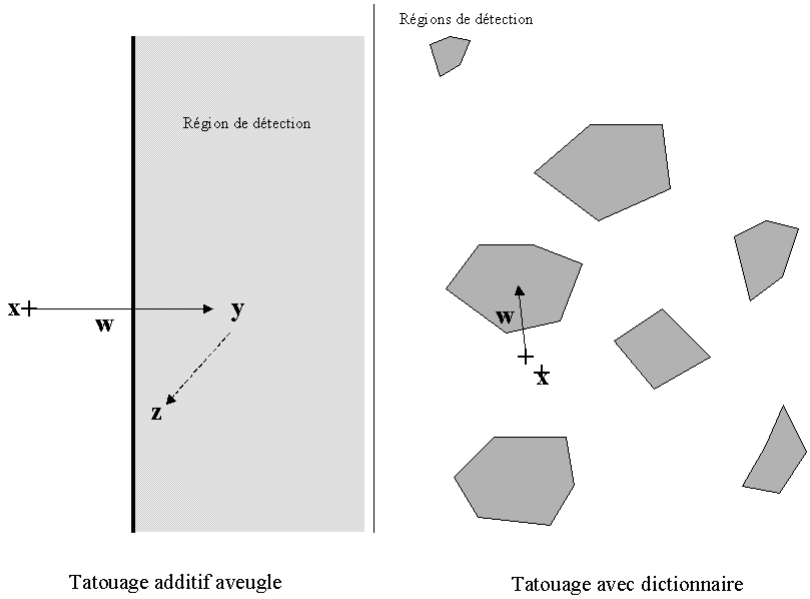

FIG. 1.4 – Stratégie d'insertion avec dictionnaire

cadre du tatouage numérique, les interférences multi-trajet n'existent pas et les MAI sont souvent négligeables devant le bruit de transmission (sauf dans le cadre du tatouage informé). Les attaques auxquelles est soumis le document sont beaucoup plus variées (cf. paragraphe 1.2.3), le pirate pouvant isoler le document de la chaîne de communication pendant un temps indéterminé avant de réintroduire une version altérée.

Modèle de tatouage multiple par étalement de spectre

Le schéma 1.5 présente le schéma général d'une chaîne de tatouage multiple par étalement de spectre. L'étape d'**insertion** consiste à :

- appliquer une transformation inversible Tr au document $\mathbf{x}$
- mettre en forme $\mathbf{m}$ en un signal redondant $\mathbf{b}$
- transformer $\mathbf{b}$ en un pseudo-bruit $\mathbf{w}$ à l'aide de $\mathbf{k}$ par l'algorithme d'étalement (*spreading*) Sp
- pondérer le message étalé par un masque psychovisuel Ψ adapté à $\mathbf{x}$
- ajouter le tatouage aux composantes transformées du document
- appliquer la transformation inverse Tr^{-1} pour obtenir un document tatoué $\mathbf{y}$ perceptuellement proche de $\mathbf{x}$.

$\mathbf{y}$ est soumis à des attaques $\mathcal{A}$ pour obtenir un document $\mathbf{z}$, toujours perceptuellement proche de $\mathbf{x}$. Le **décodage** consiste à :

- appliquer la même transformation Tr à $\mathbf{z}$
- pondérer le document tatoué par un masque psychovisuel Ψ' adapté à $\mathbf{y}$
- retrouver une estimation $\hat{\mathbf{m}}$ de $\mathbf{m}$ grâce à $\mathbf{k}$ et à l'algorithme de désétalement (*despreading*) Sp^{-1}.

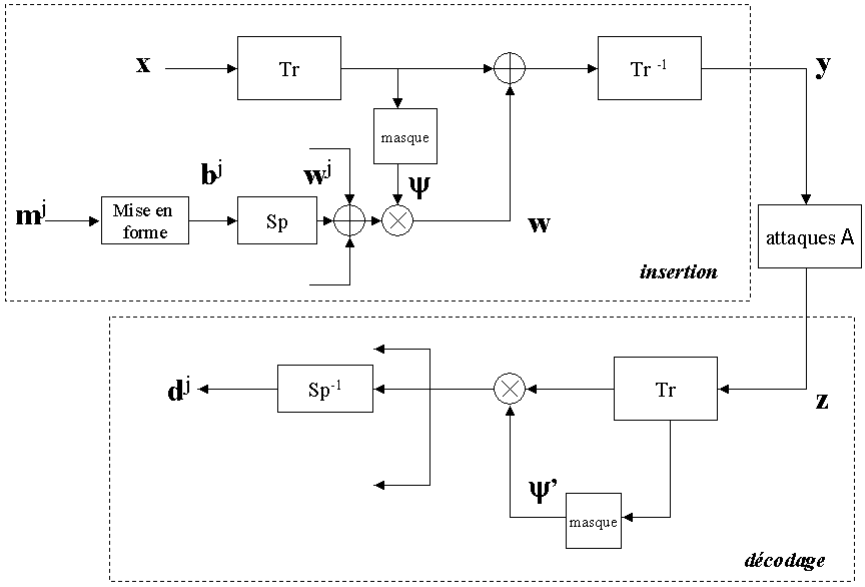

FIG. 1.5 – Chaîne de tatouage multiple par étalement de spectre

La contrainte d'imperceptibilité impose que $\mathbf{x}$ et $\mathbf{y}$ soient très proches perceptuellement. On considérera donc par la suite que $\Psi = \Psi'$. Dans le cadre du tatouage multiple, on considère J tatouages $\mathbf{m}^j$, $j = 1, \ldots, J$ étalés par Sp en $\mathbf{w}^j$, $j = 1, \ldots, J$. Dans le cas d'une insertion dans la luminance des pixels d'une image et en l'absence d'un masque ($\Psi = \psi$ Id), on a donc :

$$\mathbf{y} = \mathbf{x} + \mathbf{w} \text{ où } \mathbf{w} = \psi \sum_{j=1}^{J} \mathbf{w}^j \quad , \tag{1.5}$$

ψ est un facteur de masquage qui modère la puissance du tatouage. Les transformations inversibles utilisées en tatouage d'images seront détaillées dans le paragraphe 1.5.1.

Techniques d'accès multiple

Le principe de l'étalement de spectre (*Spread-Spectrum*) consiste à utiliser toutes les composantes fréquentielles du document. Le terme "étalement" désigne le fait de passer d'un signal possédant un spectre à bande limitée à un signal dont le spectre occupe toute la bande de fréquences. On y associe souvent le terme "blanchiment", qui consiste à passer à un signal possédant un spectre constant sur toute la bande, c'est-à-dire le spectre d'un bruit blanc. Dans l'accès multiple par division temporelle (TDMA), les messages mis en forme de chaque utilisateur sont émis les uns à la suite des autres. Dans l'accès multiple par division fréquentielle (FDMA), chaque message est modulé dans une bande de fréquence différente. Dans l'accès multiple par division par code (CDMA), chaque message est modulé par un code distinct afin que chaque message modulé occupe tout le spectre (cf. *fig.* 1.6). Lorsque la modulation est une simple multiplication par le code, on parle de *Direct-Sequence* CDMA (DS-CDMA).

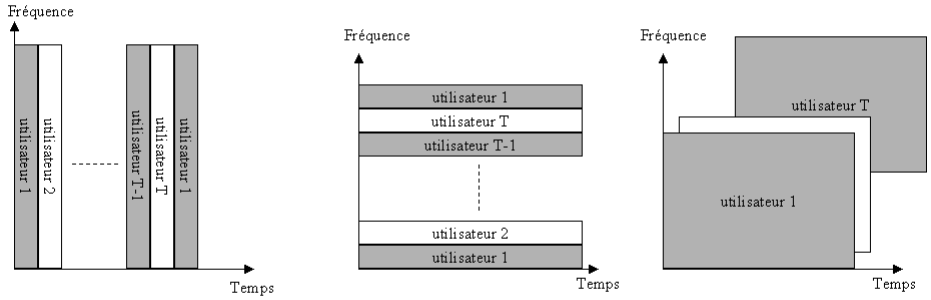

FIG. 1.6 – Principe de l'étalement TDMA, FDMA et CDMA

Tatouage par séquence directe

Tirkel et Osborne font les premiers référence à un tatouage fondé sur un pseudo-bruit, en 1993 [TRvS$^+$93]. La paternité de l'application du DS-CDMA au tatouage d'image est souvent attribuée à I.J. Cox *et al.* [CKLS96] en 1995-96. Dans un article moins connu de J. Smith *et al.* [SC96] datant de 1996, on retrouve cependant de façon plus explicite les idées de modulation de type DS-CDMA (sur toute l'image, ou par blocs pour une meilleure robustesse à la compression JPEG) ou FDMA, (équivalent d'une insertion dans le domaine fréquentiel, qui permet d'éviter la localisation spatiale mais qui est peu robuste au filtrage spatial), ainsi que d'une pré-compensation des attaques (compression JPEG). D'autres méthodes d'étalement de spectre ont été adaptées au tatouage : le TDMA repose sur une partition du signal en domaines disjoints. Dans le cas du tatouage d'image, on parlera donc plutôt de "division spatiale". La technique de FDMA par *Frequency Hopping* a notamment été utilisée dans [GDV$^+$97]. Des fonctions orthogonales de Walsh peuvent remplacer les séquences pseudo-aléatoires [MSB02].

Les algorithmes de tatouage inspirés du CDMA peuvent se ramener à un étalement par séquence directe (éventuellement dans un domaine transformé et avec une structure de code spécifique) et seront regroupés dans la suite sous le nom **DS**. L'opérateur de démodulation est utilisé sous le nom de "corrélation" en tatouage. Ses propriétés de robustesse sont primordiales dans le domaine du tatouage, il est souvent vu comme le point central de la méthode (on parle notamment de tatouage fondé sur la corrélation pour désigner les techniques de type DS). Dans le domaine des télécommunications, la redondance P est également appelée "facteur d'étalement". On module $\mathbf{b}^j$ par une

séquence pseudo-aléatoire $\mathbf{c}^j$ de moyenne nulle et de variance $\sigma^2_{\mathbf{c}} = 1$:

$$c^j_k = \pm 1, k \in \{1, \ldots, N\},$$
$$< \mathbf{c}^j, \mathbf{c}^i > = 0 \text{ pour } j \neq i, j, i \in \{1, \ldots, J\} \ ,$$

où $<,>$ est le produit scalaire. Ces J séquences orthogonales remplissent la fonction de clés secrètes ($\mathbf{k} = \{\mathbf{c}^1, \ldots, \mathbf{c}^J\}$). Le tatouage $\mathbf{w}^j = \mathbf{b}^j \mathbf{c}^j$, présente un spectre étalé. Les codes les plus courants en tatouage sont des séquences aléatoires de taille N et de distribution gaussienne ou antipodale. Au décodage, on effectue la démodulation d^j_l par corrélation :

$$d^j_l = \frac{1}{P} < \underline{\mathbf{z}}_l, \underline{\mathbf{c}}^j_l > \tag{1.6}$$

puis

$$\hat{\mathbf{m}}^j = [\text{signe}(d^j_l)]_{l \in \{(1,\ldots,L\}} \ ,$$

où signe$(x) = 1$ pour $x > 0$ et signe$(x) = -1$ pour $x < 0$. En supposant une synchronisation parfaite entre $\underline{\mathbf{z}}_l$ et $\underline{\mathbf{c}}^j_l$ ainsi qu'une orthogonalité parfaite entre les séquences, on a :

$$\hat{d}^j_l = \psi m^j_l + \frac{1}{P} < \underline{\mathbf{x}}_l + \underline{\mathbf{n}}_l, \underline{\mathbf{c}}^j_l > \ . \tag{1.7}$$

Le produit scalaire contenu dans (1.7) permet un étalement du bruit constitué par le document support. Pour un document $\mathbf{x}$ donné, lorsque P est grand, les échantillons sont supposés indépendants [HPGRN98], la seule quantité aléatoire étant $\underline{\mathbf{c}}^j_l$. Si P est suffisamment grand, le Théorème Central-Limite permet d'affirmer que

$$< \underline{\mathbf{x}}_l + \underline{\mathbf{n}}_l, \underline{\mathbf{c}}^j_l > \sim \mathcal{N}(0, \sigma^2)$$

avec

$$\sigma^2 = \sum_{k=1}^{P} (\underline{x}_{l,k} + \underline{n}_{l,k})^2 \ .$$

Or $\lim_{P \to \infty} \frac{\sigma}{P} = 0$ car le signal est borné (exemple : luminance d'une image), et pour une grande valeur de ψP, l'influence du bruit gaussien additif sur la performance de la détection est donc réduite. Cependant, lorsque ψ augmente, l'imperceptibilité du tatouage diminue et la valeur de P est limitée par la relation $N = PL$. Le rapport signal sur bruit dans (1.7) est $\psi^2 P / \sigma^2$. Le décodage par corrélation est optimal dans le cas d'une insertion additive avec un signal hôte gaussien [BBRP00]. Les performances expérimentales de DS sur une image naturelle seront étudiées dans le paragraphe 2.1.4.

Le décodeur par corrélation peut également servir de détecteur [CMB02] : en fixant un seuil de décision η, on peut décider $\hat{m}^j_l = 1$ si $d^j_l > \eta$, $\hat{m}^j_l = -1$ si $d^j_l < -\eta$ (détection du cas H_0) ou "aucun tatouage" (détection du cas H_1) si $-\eta < d^j_l < \eta$.

Tatouage à statistique de l'hôte connue

La première utilité des propriétés statistiques du document hôte en tatouage est d'estimer ce dernier à la réception. Par exemple, si $\mathbf{w}$ est stationnaire et blanc et que $\mathbf{x}$ ne l'est pas, on peut supprimer la composante non-stationnaire et non blanche de $\mathbf{z}$ avant décodage. Cette opération s'appelle "pré-blanchiment". Le pré-blanchiment au décodage consiste donc à calculer une estimation $\hat{\mathbf{x}}$ du document original à partir du

document reçu $\mathbf{z}$. Le signal préfiltré est alors $\hat{\mathbf{z}} = \mathbf{z} - \hat{\mathbf{x}}$. Au décodage, (1.2.5) devient alors :

$$\sigma^2 = \sum_{k=1}^{P}((\underline{x}_{l,k} - \underline{\hat{x}}_{l,k}) + \underline{n}_{l,k})^2 \ ,$$

et la puissance du bruit introduit par le canal est réduite (cf. paragraphe 2.4.1). Notamment, Hernandez *et al.* proposent également d'utiliser la connaissance des moments locaux du premier et second ordre pour améliorer la détection par corrélation [HPGRN98] (cf. paragraphe 1.5.5).

D'autre part, le décodeur par corrélation utilisé dans l'algorithme DS n'est optimal que pour une insertion additive et un hôte gaussien. Si cette dernière hypothèse n'est pas vérifiée, le modèle statistique du domaine d'insertion peut être utilisé pour optimiser la détection et le décodage au moyen de tests statistiques (on parle alors de méthodes avec **statistique de l'hôte connue**). Soit $f_x(x)$ la densité de probabilité dans le domaine d'insertion. La détection utilise le test de Neyman-Pearson [HPG99] :

$$\text{accepter } H_1 \text{ si} \quad \ln \frac{f_x(x|H_1)}{f_x(x|H_0)} > \eta$$

où H_1 est l'hypothèse de présence d'un tatouage et H_0 d'absence de tatouage. Au décodage, on calcule $\hat{\mathbf{m}} = \arg\max_{\mathbf{m}} \ln f_z(z|\mathbf{m})$, les mots $\mathbf{m}$ possibles étant rassemblés dans un dictionnaire. On utilise ici le principe du Maximum de Vraisemblance (tous les messages sont supposés avoir la même probabilité *a priori*). C'est également l'équivalent d'un test d'hypothèse à $|\mathcal{M}|$ hypothèses, suivant le critère du maximum *a posteriori* (MAP) [JHM01]. L'utilisation du détecteur optimal et du décodeur optimal conduit à une nette amélioration des performances (cf. paragraphe 2.4.2).

L'utilisation de la statistique de l'hôte au bénéfice du tatouage a été également envisagée pour les techniques de catégorisation aléatoire (cf. paragraphe 1.3.2).

Tatouage et codage canal

L'analogie entre tatouage et télécommunications conduit à envisager l'utilisation de codes correcteurs d'erreur en amont de l'étalement, à la place de la mise en forme : c'est le domaine du codage canal. Au lieu de $\mathbf{b}$, on soumet donc à Sp une succession de mots de code de taille N et appartenant à un dictionnaire $\mathcal{M}$. La technique de mise en forme avant étalement utilisée dans DS peut être elle-même vue comme un code correcteur rudimentaire, dit **code par répétition** ou "technique de diversité". De manière générale, les spécificités du tatouage à insertion aveugle (rapport signal à bruit très faible) font que les codes correcteurs d'erreur classiques sont moins efficaces que dans le domaine des télécommunications. D'autres notions de codage (codage aléatoire, *binning*), plus liées au codage source ou codage conjoint source/canal, sont utilisées avec plus de succès dans les méthodes de tatouage informé (cf. paragraphe 1.2.6). Les techniques proposées dans ce rapport de thèse peuvent être combinées avec des codes correcteurs d'erreur au même titre que DS, pour le même gain de performance.

1.2.6 Tatouage informé

Le tatouage informé s'appuie sur deux piliers : les principes du codage informé et de l'insertion informée présentés par Cox et Miller et le schéma de Costa. Les algorithmes pratiques inspirés de ces principes sont souvent des techniques de tatouage substitutif avec dictionnaire.

Principe du tatouage informé

Le tatouage informé a été proposé en 1999 par I.J. Cox et M. Miller [CMM99], à partir de l'observation selon laquelle le signal hôte, considéré jusqu'ici comme une source de bruit de transmission, est parfaitement connu à l'insertion. Ils proposent deux cadres à l'élaboration de techniques de tatouage informé : le codage informé et l'insertion informée [MDC04][MK05].

Le **codage informé** consiste à construire des tatouages $\mathbf{w}$ dépendants du signal hôte $\mathbf{x}$. Cette définition englobe les techniques classiques de masquage perceptuel [CMB02]. Cependant, ce principe suggère de calculer $\mathbf{w}$ directement à partir de $\mathbf{x}$, plutôt que de construire un tatouage pour ensuite l'adapter perceptuellement à $\mathbf{x}$. Cela correspond également à construire un dictionnaire de mots de codes. Un message est associé à un ou plusieurs mots de code, le mot inséré étant déterminé par le signal hôte.

L'**insertion informée** repose en outre sur la connaissance à l'insertion de la structure du détecteur et de ses régions de détection correspondantes, tenant compte d'éventuelles attaques. La démarche consiste à adapter $\mathbf{w}$ au signal hôte, selon diverses contraintes d'imperceptibilité, de détection, de robustesse. Par exemple, dans ce dernier cas, la contrainte est que $\mathbf{z}$ soit contenu dans la région de détection. Certains détecteurs utilisent plusieurs régions de détection pour le même symbole. Dans ce cas, l'insertion informée suggère d'insérer le tatouage dans la région qui introduit le moins de distorsion. L'insertion aveugle classique consiste à maximiser la Corrélation Linéaire (MLC) à distorsion constante, ce qui selon l'hôte ne permet pas toujours de détecter le tatouage. Trois stratégies simples d'insertion informée sont présentées dans [MCB00] : Maximisation de la Robustesse (MR) à distorsion constante, Robustesse Constante (CR) pour optimiser la distorsion, la capacité ou la performance, et Maximisation du Coefficient de Corrélation (MCC) à distorsion constante pour garantir la détection lorsqu'elle est possible au détriment de la robustesse.

La *fig.* 1.7 donne une interprétation géométrique des stratégies de tatouage informé. Les documents tatoués respectant la contrainte d'imperceptibilité sont représentés par un disque de centre $\mathbf{x}$. L'insertion a pour but de créer un document tatoué appartenant à une région de détection. Ici, la région de détection est représentée par un cône, qui convient au décodage par corrélation normalisée. En effet, dans ce cas, le produit scalaire à la détection peut se ramener à une valeur angulaire : $\frac{<\mathbf{a},\mathbf{b}>}{\|\mathbf{a}\|\|\mathbf{b}\|} = \cos\theta$ [CMB02]. Plus $\mathbf{y}$ est situé profondément au cœur de la région de détection (ici, vers la droite), plus la robustesse est grande. Si l'on arrive à exprimer une mesure de la robustesse, on peut alors tracer un contour de robustesse constante. Dans le cas d'une corrélation normalisée, il s'agit d'une hyperbole tangente au cône[1]. L'insertion aveugle consiste à insérer un tatouage parallèle à l'axe du cône, pour maximiser le coefficient de détection en l'absence de bruit de l'hôte (*Maximum Linear Correlation*, MLC). On voit que la région de distorsion acceptable et la région de détection ne se coupent pas toujours sur cet axe. La maximisation de la robustesse (MR) à distorsion constante consiste à choisir le point du disque de distorsion acceptable situé la plus profondément dans la région de détection. La robustesse constante (CR) consiste à choisir le point d'un contour de robustesse constante le plus proche de $\mathbf{x}$. Sur cette figure, CR ne peut pas respecter la contrainte d'imperceptibilité.

[1] $\tau_{\text{nc}} = \frac{<\mathbf{z},\mathbf{c}>}{\|\mathbf{z}\|\|\mathbf{c}\|} \simeq \frac{<\mathbf{y},\mathbf{c}>}{\sqrt{\|\mathbf{y}\|^2+\|\mathbf{n}\|^2}\|\mathbf{c}\|} \Rightarrow \sigma_{\mathbf{n}}^2 \simeq \left(\frac{<\mathbf{y},\mathbf{c}>}{\tau_{\text{nc}}\|\mathbf{c}\|}\right)^2 - \|\mathbf{y}\|^2$

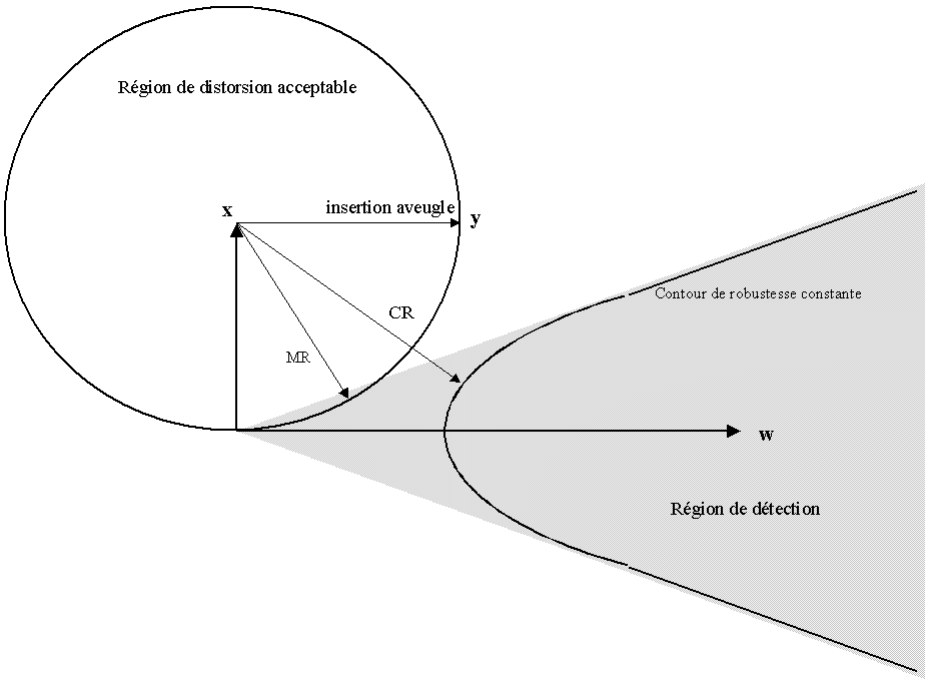

FIG. 1.7 – Stratégies d'insertion informée

Implantations du principe de Cox et Miller

Les algorithmes de tatouage informé sont plus complexe à implanter que les algorithmes de tatouage aveugles. En effet, ils imposent une connaissance du détecteur, de l'impact d'éventuelles attaques sur les zones de détection, et surtout d'une technique pour faire passer un mot de code dans une zone de détection. L'implantation la plus simple, fondée sur l'étalement de spectre et sous-optimale, est appelée "étalement de spectre amélioré" (ISS). Elle sera détaillée dans le paragraphe 1.3.1. D'autres techniques reposent sur la modification de codes correcteurs d'erreurs pour construire un dictionnaire de mots de codes dépendant de $\mathbf{x}$. L'insertion informée dépend ensuite du décodeur associé au code correcteur d'erreurs. Contrairement aux méthodes quantificatives (cf. paragraphe 1.3.2), ces techniques de tatouage informé ne sont pas particulièrement sensibles aux attaques valumétriques car elles utilisent un décodage fondé sur la corrélation.

L'algorithme de Miller, Doërr et Cox [MDC04], utilise un code treillis. Un code treillis consiste en un graphe de nœuds reliés par des arcs, comprenant L niveaux et 8 nœuds par niveau, d'où 8^L chemins possibles. D'un nœud ne partent que 2 arcs indexés par m_l (0 ou 1) : le treillis permet donc de sélectionner 2^L chemins dans l'arbre (1 chemin par message possible). 2 messages différant d'1 seul bit diffèrent de 4 arcs une fois codés, ce qui améliore la robustesse au bruit et permet d'effectuer un décodage souple itératif (algorithme de Viterbi). Dans l'algorithme de tatouage, on modifie le treillis afin que le nombre d'arcs partant d'un nœud soit supérieur à 2 (ex : 4) et donc que plusieurs chemins correspondent au même mot. Le chemin choisi parmi ce sous-dictionnaire correspondant à $\mathbf{m}$ est celui qui possède la plus grande corrélation avec $\mathbf{x}$. Il est déterminé à l'aide d'un algorithme de décodage de Viterbi appliqué sur le

sous-treillis correspondant à **m**. L'insertion informée procède ensuite de manière itérative et sous-optimale pour adapter ce mot de code au signal. On se donne une région de détection R_g à atteindre (dépendant de contraintes). On l'atteint à partir du mot de code treillis $\in R_0$ en remplaçant le tatouage à chaque étape i par un tatouage situé derrière la frontière entre les zones de détection R_g et R_i jusqu'à atteindre R_g. Le point atteint n'est pas forcément le point optimal dans R_g. La combinaison des deux techniques offre d'excellents résultats en termes de capacité et de robustesse, l'un semblant profiter à l'autre. L'imperceptibilité est mesurée par la distance de Watson (cf. paragraphe 1.5.3) et peut être améliorée par un masque fondé lui aussi sur le modèle de Watson.

Cette technique peut être reliée à celle de Chou et Ramchandran qui modifient pour leur part des codes correcteurs fondés sur des codes blocs syndromes [CPR99] ou des codes treillis turbo-codés [CPR01]. Le Guelvouit et Pateux [GP03] ont quant à eux appliqué ce principe à des codes poinçonnés convolutifs et à un décodage souple par treillis, sous la terminologie de *Wide Sense Spread Spectrum*. Enfin, Abrardo et Barni [AB04] proposent d'utiliser des codes orthogonaux (cas particulier de codes blocs), pouvant être insérés plus efficacement. Soit un dictionnaire $\mathcal{U}$, matrice carrée unitaire de N_l mots de codes de taille N_l, orthogonaux entre eux. On divise $\mathcal{U}$ en L sous-dictionnaires. Dans chacun de ces sous-dictionnaires, on choisit comme code d'étalement celui qui présente la plus grande corrélation avec la portion de l'image considérée. En pratique, on utilise des codes de Gold à la place des codes orthogonaux pour leur meilleures propriétés de correction des erreurs, ainsi qu'un turbo-codage. Cette technique est plus simple que les précédentes. Elle se rapproche de celle de Miller *et al.*, sans la complexité supplémentaire introduite par le treillis.

D'autres techniques d'insertion informée sont évoquées dans [DFHS03]. PEAK (*Peaking DS*) est une version simplifiée de LISS issue de [CMB02], où $\lambda = 1$ (encore appelée pré-annulation de l'image). ZATT (*Zero Attraction*) est un détecteur du N-ième ordre proposé par Furon *et al.* et utilisant des projections secrètes. Furon *et al.* ont proposé une autre technique d'insertion et de détection (mais pas de décodage) informée appelée JANIS [FMSH02][Fur02]. Enfin, l'étalement de spectre amélioré est une adaptation particulièrement simple de la technique DS à l'insertion informée par Maximisation de la Robustesse. Cette technique est détaillée dans le paragraphe 1.3.1.

Principe de Costa

La capacité d'une technique de tatouage représente la quantité d'information maximale qui peut en théorie être cachée dans un document. Elle dépend du modèle statistique du document, des contraintes d'imperceptibilité imposées au tatoueur et au pirate, et de l'information disponible lors de l'insertion, des attaques, et du décodage. La capacité est indépendante de l'algorithme de tatouage. L'un des objectifs de la conception d'un algorithme de tatouage est d'atteindre en pratique la capacité tout en respectant les contraintes d'imperceptibilité et de robustesse. On parle alors de débit accessible (*achievable rate*).

Capacité d'un canal sans information adjacente : la capacité est exprimée en bit/échantillon. Soit H la fonction entropie de la variable aléatoire X :

$$H(X) = -\sum_{x} \mathrm{p}(x) \log_2(\mathrm{p}(x))$$

et I l'information mutuelle (ou transinformation) de l'entrée X d'un canal et de sa sortie Y :

$$\mathrm{I}(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$$

Alors on appelle capacité de ce canal

$$C = \max_{\mathrm{p}(x)} \mathrm{I}(X;Y)$$

La capacité d'un canal transmettant un signal gaussien $X \sim \mathcal{N}(0, \sigma_X^2)$ et perturbé par un bruit gaussien $N \sim \mathcal{N}(0, \sigma_N^2)$, appelé canal gaussien, est

$$C_{\mathrm{GC}} = \frac{1}{2} \log_2 \left[1 + \frac{\sigma_X^2}{\sigma_N^2} \right]$$

C'est ce calcul de capacité qui prévaut pour les schémas de tatouage à insertion aveugle classiques du type étalement de spectre, X étant le message à transmettre, Y le document tatoué, et N recouvrant à la fois le bruit dû aux attaques et dû au document hôte. On a fait ici l'hypothèse d'un canal de tatouage gaussien. Les pixels d'une image ne suivant pas une loi gaussienne et étant corrélés entre eux, certains calculs de capacité modélisent une transformation de l'image comme un mélange de lois. Certains espaces d'insertion sont modélisés de façon plus fine. Une étude spécifique est par exemple possible pour une insertion multiplicative dans le domaine spectral [BBRP00].

Canal avec information adjacente : on appelle canal avec information de bord (ou information adjacente) un canal gaussien perturbé par le bruit N inconnu et par un signal S connu. Gel'fand et Pinsker [GP80] ont montré que la capacité de ce canal est

$$C = \max_{\mathrm{p}(u,x|s),\mathrm{p}(s),\mathrm{p}(y|x,s)} \mathrm{I}(U;Y) - \mathrm{I}(U;S)$$

Cette capacité peut être atteinte via l'utilisation d'un signal auxiliaire U connu à l'émission et à la réception, qui fait office de mot de code choisi dans un dictionnaire. Ce dictionnaire $\mathcal{U}$ est divisé en $|\mathcal{M}|$ ensembles $\mathcal{U}_{\mathbf{m}}$ (où $|\mathcal{M}|$ est le nombre de mots messages possibles). Ces ensembles, aussi appelés *bins*, sont choisis de sorte à maximiser la distance entre les mots de code qu'ils contiennent. A l'encodage, U est choisi tel que $E[(U - \alpha S)S] = 0$. On dit que U est "typique conjointement" avec S. Au décodage, on identifie le *bin* qui contient le mot de code typique conjointement avec Y. Par opposition au codage aléatoire (cas où aucune information de bord n'est utilisée), cette technique est appelée **catégorisation aléatoire** (pour *random binning*, aussi traduisible par "regroupement", "partitionnement" ou "groupement par classe").

Schéma de Costa : les résultats précédents ont été étendus dans un article de M. Costa publié en 1983 et intitulé *"Writing on dirty paper"* [Cos83]. Cet article a été repris 17 ans plus tard par la communauté du tatouage et sert de base à toute une branche du tatouage numérique. Costa étudie le cas où S est gaussien ($S \sim \mathcal{N}(0, \sigma_S^2)$) et en utilisant la distance des moindres carrés. Dans un canal gaussien traditionnel, on aurait $C = \frac{1}{2} \log_2 \left[1 + \frac{\sigma_X^2}{\sigma_N^2 + \sigma_S^2} \right]$. Costa a montré que l'on pouvait atteindre la capacité

$$C_{\mathrm{GCSI}} = \frac{1}{2} \log_2 \left[1 + \frac{\sigma_X^2}{\sigma_N^2} \right]$$

De plus, Costa a construit une méthode pour atteindre asymptotiquement cette capacité. Il propose de transmettre $X = U - \alpha S$, avec $\alpha_{\mathrm{optimal}} = \frac{\sigma_X^2}{\sigma_N^2 + \sigma_S^2}$. Dans le schéma de

Costa optimal, on construit un dictionnaire $\mathcal{U}$ constitué de valeurs tirées aléatoirement selon $\mathcal{N}(0, \sigma_X^2 + \alpha^2\sigma_S^2)$. La taille de $\mathcal{U}$ croît exponentiellement avec L et en pratique on doit utiliser des dictionnaires sous-optimaux structurés (cf. paragraphe 1.3.2). Cette technique est appelée "insertion avec information de bord" ou *dirty paper coding*.

Dans le cadre du tatouage, on peut considérer que $X = \mathbf{w}$ est le tatouage, $N = \mathbf{n}$ le bruit introduit par les attaques, $S = \mathbf{x}$ le bruit introduit par le document support (connu à l'insertion) et $U = \mathbf{u}$ un signal auxiliaire. Par exemple, dans les techniques quantificatives, $\mathbf{u}$ est une version quantifiée du document, cf. partie 1.3.2. L'utilisation d'une insertion avec information de bord permet d'obtenir la même performances en tatouage aveugle qu'en tatouage non aveugle. La formule de Costa a été étendue au cas de toute variable aléatoire S à puissance finie : la capacité d'un canal de tatouage est indépendante de la variance de l'hôte [Mou02]. Ces résultats incitent à appliquer les résultats de Costa à un système de tatouage pratique. La terminologie *dirty paper coding* tend à céder la place au terme *random binning* ([TVKP05],[ZD05],[MK05]).

Par construction, les techniques de tatouage par catégorisation aléatoire sont des techniques de tatouage informé : $\mathbf{u}$ est sélectionné dans $\mathcal{U}_\mathbf{m}$ en fonction de $\mathbf{x}$ afin de maximiser la capacité (choisie ici comme critère de performance, donc comme stratégie d'insertion). L'implantation la plus courante du schéma de Costa, faisant appel à la quantification, a d'ailleurs été proposée par Chou *et al.* [CPR99] et Chen et Wornell [CW01] au même moment que le tatouage informé l'était par Cox et Miller.

Codes de tatouages pratiques : le tatouage étant désormais modélisé par une technique de codage dans un canal avec information adjacente, la conception d'un algorithme de tatouage peut être vue comme la construction de codes optimaux [MK05]. Notamment, la construction d'un code revient à un remplissement compact (ou *sphere packing*) de l'espace de tatouage par des zones de détection. Dans la technique DS et son adaptation à l'insertion informée LISS (cf. paragraphe 1.3.1), on utilise seulement deux mots de codes binaires antipodaux : il ne s'agit pas de codage informé. Les applications du principe de Cox et Miller utilisant des mots de code à énergie égale et un décodage impliquant une corrélation, comme celles de Miller, Doërr et Cox et celle de Abrardo et Barni, peuvent être considérées comme des techniques de catégorisation aléatoire à codes sphériques [PGMBA05] : on peut montrer que l'ensemble des mots de codes est situé sur la surface d'une hypersphère de rayon 1, et les régions de détection sont des hypercônes. Grâce à cette propriété, ces techniques sont invariantes aux attaques de gain : un changement d'amplitude du signal ne le fait pas sortir d'un cône de détection (cf. *fig.* 1.8). Les codes de tatouage les plus utilisés en pratique sont les grilles de quantification[2], où le centre des régions de détection suit un arrangement régulier, invariant par translation. Ces techniques ne sont pas intrinsèquement invariantes à l'attaque de gain. Elles seront développées en détail dans le paragraphe 1.3.2.

1.2.7 Tatouage substitutif

Les techniques de tatouage substitutif, très présentes dans les débuts du tatouage numérique, fournissent les algorithmes les plus atypiques car moins liés aux domaines des télécommunications et du codage source.

[2]Remarque : on distinguera les mots anglais *trellis*, ou treillis, qui désigne un arbre et est utilisé dans les codes du même nom, et *lattice*, également traduisible par treillis, mais qui désigne une grille régulière. On parlera donc de grilles de quantification.

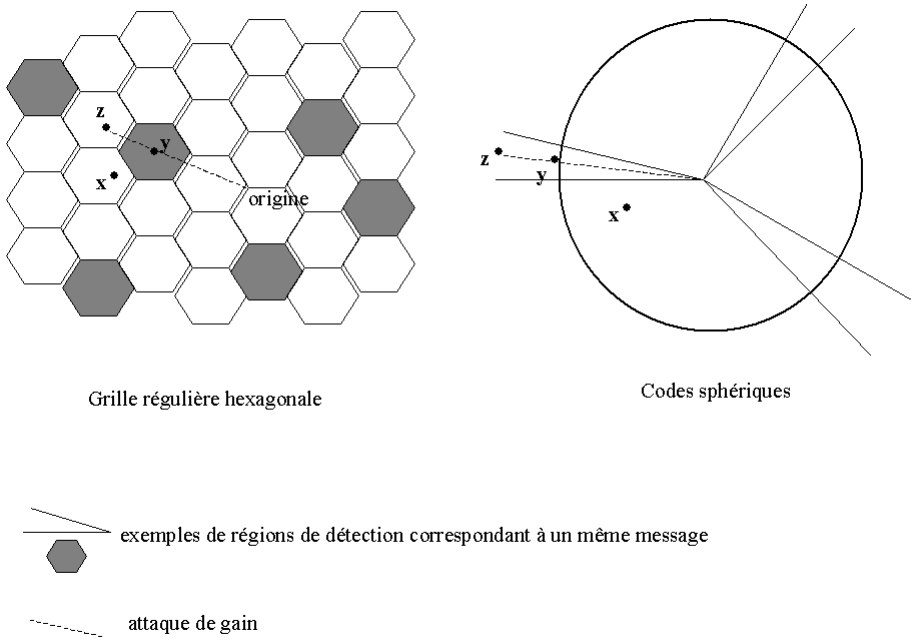

FIG. 1.8 – Exemples de codes de tatouage sphériques et hexagonaux en 2D

Algorithmes substitutifs assimilables aux techniques quantificatives : les premières techniques de tatouage (1993) remplaçaient directement les bits de poids faibles (LSB pour *Least Significant Bit*) de **x** par **b** [TRvS$^+$93]. Cependant, cette méthode est très sensible au bruit, additif en particulier. Son champ d'application se limite donc souvent à la stéganographie ou au tatouage fragile [YM97]. La technique LSB est liée au tatouage quantificatif (cf. paragraphe 1.3.2).

Algorithmes substitutifs assimilables à l'étalement de spectre : dans la technique du *patchwork* (1996) [BGML96], on sélectionne un ensemble de paires d'échantillons du signal original d'amplitude $(a_i, b_i)_{i=1,\dots,P}$. Le codage consiste à augmenter a_i tout en diminuant b_i, d'une même valeur pré-définie. Le détecteur consiste à seuiller $\sum_i a_i - \sum_i b_i$. De nombreuses itérations, effectuées sur des ensembles disjoints et en générant aléatoirement les paires à chaque itération, permettent de diminuer la variance de la statistique de détection, et donc d'améliorer ses performances. Les ensembles de points peuvent prendre des formes particulières aux contours variés (les *patchs*), couvrant différentes fréquences. Un schéma très proche du *patchwork*, appelé *signature casting*, a été développé simultanément [NP98]. Il utilise une seule itération et, en l'absence d'attaque, le seuil au décodage est déterminé statistiquement grâce à un test d'hypothèses utilisant $\sigma^2_{\mathbf{x}}$. La détection du *patchwork* peut s'interpréter par un calcul de corrélation entre l'image tatouée et la séquence ternaire valant 1 pour les points correspondant à a_i, -1 pour ceux correspondant à b_i et 0 pour les points non modifiés. Les séquences correspondant à des ensembles disjoints ou à des itérations différentes sont orthogonales. Ces méthodes seront donc assimilées dans ce qui suit à la méthode DS.

Algorithmes substitutifs avec contraintes (exemples en tatouage d'image) : Burgett, Koch et Zao (1998) [BKZ98] ont élaboré une méthode d'insertion substitutive particulièrement populaire correspondant aux algorithmes substitutifs avec contraintes. Ils proposent de construire un "vecteur caractéristique" contenant le résultat de la comparaison de N paires d'échantillons choisis dans le domaine de la DCT d'une image. Les échantillons sont ensuite permutés au sein de chaque paire afin que le vecteur caractéristique contienne $\mathbf{b}$.

Plusieurs méthodes de tatouage sont inspirées de la compression fractale. La compression fractale recherche des similarités au sein d'une image : pour un bloc original de $\mathbf{x}$ donné Rb_j (*Range Block*), on recherche dans l'image le bloc Db_j (*Domain Block*) et la transformation affine tels que $T_j(Db_j)$ soit le plus proche de Rb_j, dans le but de ne stocker que T_j et Db_j. La transformation globale $T = \{T_j\}$ doit être contractante et avoir pour attracteur $\mathbf{x}$. Puate et Jordan [PJ96] proposent de tatouer le "code fractal" (T_j, Db_j) de l'image en limitant les domaines de recherche de Db_j à deux ensembles $\mathcal{S}_{-1}$ et $\mathcal{S}_{+1}$, selon le bit inséré. La méthode d'insertion de similarités [BCD98] consiste pour sa part à rajouter artificiellement des similarités parfaites entre des blocs de l'image. Le code fractal est donc complètement modifié. Ces deux méthodes s'appuient donc sur une substitution du code fractal.

L'élaboration de la méthode de tatouage substitutif est souvent motivée par la robustesse à une attaque particulière. Face aux attaques géométriques, la substitution s'appliquera à des caractéristiques telles que les points saillants, situés dans les régions de plus haute énergie. Les positions relatives de ces points étant presque invariantes par transformations géométriques, le tatouage est plus robuste à ces attaques. Maes *et al.* [MO98] proposent de les faire coïncider avec un réseau dense de droites représentant leurs positions relatives, par substitution par des points situés sur ce réseau.

1.2.8 Techniques de tatouage inspirées d'autres disciplines

Le tatouage numérique se situe à la convergence de nombreuses disciplines : télécommunications, traitement de l'image, théorie de l'information, codage source et canal, cryptographie, statistiques, classification, théorie des jeux... Le problème est notamment particulièrement adapté à la classification et à l'optimisation sous contraintes.

Prise en compte des contraintes du problème et théorie des jeux : la capacité est utilisée dans une modélisation du tatouage par la théorie des jeux [MO03]. Le tatoueur cherche à maximiser la capacité pour une énergie d'insertion donnée, alors que le pirate cherche à la minimiser pour une énergie d'attaque donnée. Il s'agit donc d'un jeu dit du $\max\min$. Cette modélisation apporte des indications sur les stratégies d'insertion à privilégier dans l'élaboration d'algorithmes pratiques.

Utilisation de techniques de classification : les techniques de classification sont avant tout utilisées pour choisir des zones propices au tatouage. C'est le cas des régions à tatouer d'une image dans [BSNO99] (*k-means, fuzzy c-means*) et des points saillants en couleur (alors situés sur les centroïdes de *clusters* de l'algorithme *k-means*) dans [CSST01]. Les méthodes de classification sont également utiles à la resynchronisation, lorsque le schéma de tatouage n'est pas aveugle. Diverses méthodes d'apprentissage sont enfin utilisées pour améliorer la détection et/ou sélectionner les meilleurs coefficients d'insertion : réseaux de neurones [YTL01], algorithmes génétiques [HW00][SHWP04][KA05] et *Support Vector Machines* (SVM) [FSL04].

Tatouage et optimisation sous contraintes : le tatouage informé peut être considéré comme un problème d'optimisation sous différentes contraintes : imperceptibilité, détection, robustesse à une attaque donnée (ex : compression), fragilité à d'autres attaques (ex : attaques géométriques). . . Notamment, Pereira *et al.* [PVP01] proposent de considérer l'insertion comme un problème de programmation linéaire : on veut maximiser la force du tatouage sous des contraintes de distorsion. L'optimisation repose sur l'algorithme du simplexe. Altun *et al.* [ASB05][ASB06] proposent de simplifier le problème de l'insertion en représentant l'ensemble des tatouages satisfaisant une contrainte par un ensemble. La solution du problème d'optimisation est alors tout tatouage situé dans l'intersection de ces ensembles. L'intersection est plus facile à calculer lorsque les ensembles sont convexes. Notamment, les Projections sur des Ensembles Convexes (POSC) sont utilisées pour optimiser la capacité sous contraintes de détectabilité, fidélité visuelle et robustesse à la compression. La détection rentrant en compte dans les contraintes, il s'agit d'une méthode de tatouage informé, qui a pour principal défaut un temps de calcul très important.

Tatouage orienté contenu ou objet : l'extraction du contenu du document interagit de plusieurs manières avec le tatouage. Le contenu d'une image est par exemple constitué d'objets, issus d'une extraction de contours. Une technique de tatouage traditionnelle peut alors insérer dans chaque objet sa propre description [BKMS01]. Dans [KL02], la technique de détection utilise une classification des zones de l'image selon leur texture. D'autres techniques utilisent la forme des objets. Une approche particulièrement intéressante est celle de [BM01]. Elle anticipe l'utilisation de la norme MPEG-4, dans laquelle les objets vidéos sont décrits selon leur texture et leur forme. Leur échelle et leur position temporelle peuvent être manipulées. Les techniques de tatouage doivent donc d'une part agir sur les objets eux-mêmes, et surtout être robustes à ces manipulations (translation, rotation, changement d'échelle, déplacement temporel), par exemple avec des signatures orientables tenant compte de la géométrie de l'objet.

1.3 Techniques pratiques de tatouage informé

Le principe du tatouage informé a été présenté dans le paragraphe 1.2.6. Dans ce paragraphe, on présente plus précisément deux types d'algorithmes pratiques de tatouage informé : l'étalement de spectre amélioré et les techniques quantificatives. Dans le chapitre 2, des adaptations de LISS et ST-SCS seront proposées. Dans les chapitres 2, 3 et 4, les performances des techniques proposées seront comparées avec celles de LISS, SCS, ST-SCS et RDM.

1.3.1 Étalement de spectre amélioré

Malvar et Florêncio ont proposé une technique appelée étalement de spectre amélioré (ou *Improved Spread Spectrum*, ISS) [MF03]. Elle consiste à exploiter la connaissance de l'hôte à l'insertion pour moduler l'énergie du tatouage. Cette modulation permet de compenser les interférences dues à l'hôte. Il s'agit donc d'une technique d'insertion informée qui exploite la stratégie de Maximisation de la Robustesse (MR) (cf. paragraphe 1.2.6), la distorsion étant fixée à sa valeur habituelle dans l'algorithme DS classique. La méthode est censée offrir les mêmes performances que la QIM (cf. paragraphe 1.3.2) tout en restant dans le cadre de l'étalement de spectre. C'est le cas face à l'attaque AWGN pour un faible WNR. Si le bruit est faible (WNR grand), ST-SCS est

meilleur [PFPGV06]. La robustesse de l'ISS face à d'autres attaques est cependant à étudier expérimentalement. Delhumeau *et al.* [DFHS03] évoquent diverses techniques d'amélioration de l'étalement, et en concluent que l'ISS est le plus performant pour le problème de la détection.

Étalement de spectre amélioré linéaire (LISS) : dans sa version la plus simple, on insère :

$$\mathbf{w}' = (\alpha\mathbf{b} - \lambda\frac{< \mathbf{x}, \mathbf{c} >}{\|\mathbf{c}\|^2})\mathbf{c} \tag{1.8}$$

au lieu de $\mathbf{w} = \alpha\mathbf{bc}$ dans le schéma (1.4). La distorsion *a posteriori* est pour un bit d'information

$$D_w = \mathrm{E}[|\alpha b_k - \lambda\frac{< \mathbf{x}, \mathbf{c} >}{\|\mathbf{c}\|^2}|^2\mathbf{c}^2] = \left(\alpha^2 + \frac{\lambda^2\sigma_{\mathbf{x}}^2}{P\sigma_{\mathbf{c}}^2}\right)\sigma_{\mathbf{c}}^2$$

λ contrôle l'erreur introduite par le document support lors du décodage. Un compromis doit être trouvé entre α et λ pour minimiser à la fois l'erreur de décodage et la déformation. Ainsi, pour conserver la même distorsion $\sigma_{\mathbf{c}}^2$ que dans le schéma (1.4), on pose :

$$\alpha = \sqrt{\frac{P\sigma_{\mathbf{c}}^2 - \lambda^2\sigma_{\mathbf{x}}^2}{P\sigma_{\mathbf{c}}^2}} \tag{1.9}$$

Alors

$$\mathrm{TEB} = Q\left(\sqrt{\frac{P\sigma_{\mathbf{c}}^2 - \lambda^2\sigma_{\mathbf{x}}^2}{(1-\lambda)^2\sigma_{\mathbf{x}}^2 + \sigma_{\mathbf{n}}^2}}\right). \tag{1.10}$$

Donc le TEB est optimal pour λ_{opt} tel que $\partial\mathrm{TEB}/\partial\lambda = 0$, c'est-à-dire :

$$\lambda_{\text{opt}} = \frac{1}{2}\left((1 + \frac{\sigma_{\mathbf{n}}^2}{\sigma_{\mathbf{x}}^2} + \frac{P\sigma_{\mathbf{c}}^2}{\sigma_{\mathbf{x}}^2}) - \sqrt{\left(1 + \frac{\sigma_{\mathbf{n}}^2}{\sigma_{\mathbf{x}}^2} + \frac{P\sigma_{\mathbf{c}}^2}{\sigma_{\mathbf{x}}^2}\right)^2 - 4\frac{P\sigma_{\mathbf{c}}^2}{\sigma_{\mathbf{x}}^2}}\right) \tag{1.11}$$

Pour P grand, $\lambda_{\text{opt}} \to 1$. Dans la version non linéaire, λ n'est plus constant et varie avec $\mathbf{x}$ et $\mathbf{m}$ pour de meilleurs compromis entre TEB et DWR.

On peut généraliser le LISS [Bru03] :

$$\mathbf{w}' = g(\mathbf{x}, \mathbf{m})$$

où g est une fonction d'étalement tenant compte de l'information de bord. g est construite pour maximiser le TEB, elle peut donc inclure une phase de préfiltrage. Comme pour l'étalement de spectre classique, on peut pondérer le tatouage par un masque psychovisuel Ψ [CT03] :

$$\mathbf{w}' = \Psi(\alpha\mathbf{b} - \lambda\frac{< \mathbf{x}, \Psi\mathbf{c} >}{\|\Psi\mathbf{c}\|^2})\mathbf{c}$$

Ainsi, on a au décodage :

$$< \mathbf{z}, \Psi\mathbf{c} >= \alpha\mathbf{b}\Psi^2 + (1-\lambda) < \mathbf{x}, \Psi\mathbf{c} > + < \mathbf{b}, \Psi\mathbf{c} >$$

Limite du rejet des interférences de l'hôte : $0 < \lambda < 1$. En l'absence de bruit, LISS rejette les interférences de l'hôte si et seulement si $\lambda = 1$. Dans ce cas, le dénominateur $(1-\lambda)^2\sigma_{\mathbf{x}}^2 + \sigma_{\mathbf{n}}^2 = 0$ et le TEB=0. Mais alors comme $\alpha = \sqrt{\frac{P\sigma_{\mathbf{c}}^2 - \lambda^2\sigma_{\mathbf{x}}^2}{P\sigma_{\mathbf{c}}^2}}$, on doit avoir $\lambda^2 < \frac{P\sigma_{\mathbf{c}}^2}{\sigma_{\mathbf{x}}^2}$, ou encore $P > \frac{\sigma_{\mathbf{x}}^2}{\sigma_{\mathbf{c}}^2}$. Dans LISS, la part de la puissance du tatouage consacrée à la suppression des interférences de $\mathbf{x}$ ne peut être inférieure à $\sigma_{\mathbf{x}}^2/P$, ce qui limite DWR. Le rejet des interférences n'est donc possible que si P est suffisamment grand ou si DWR est suffisamment faible. Par exemple, si DWR=28, pour l'image Lena, si $L > 327$, le TEB est non nul. L'utilisation du préfiltrage de Wiener (cf. paragraphe 1.5.5) est alors obligatoire pour avoir un TEB proche de celui obtenu par les méthodes quantificatives. Dans ce cas, $\sigma_{\mathbf{x}}^2$ devient $\sigma_{\mathbf{x}-\hat{\mathbf{x}}}^2$ et la limite inférieure de rejet est $P > \frac{\sigma_{\mathbf{x}-\hat{\mathbf{x}}}^2}{\sigma_{\mathbf{c}}^2}$ (par exemple $L < 17656$ pour Lena et DWR=28 dB).

LISS et scénarios d'attaques : λ_{opt} ne peut être calculé que dans le scenario de l'attaque AWGN avec $\sigma_{\mathbf{n}}^2$ connu. Pour d'autre attaques connues, on doit essayer d'estimer leur variance équivalente en AWGN. Si l'attaque est inconnue, on peut fixer un $\sigma_{\mathbf{n}}^2$ maximum. Dans les autres cas, un bon choix de λ est $\lambda = 1$, ce qui correspond à ne pas compenser les distorsions (cf. 1.3.2). La figure 4 de [MF03] tend à montrer que λ_{opt} est très proche de 1 (du moins, il n'y a pas grande perte de TEB pour $\lambda = 1$). Ceci est vrai pour des valeurs pratiques de WNR et P. Pour les valeurs de L utilisées ici ($L < 1024$), $0.95 < \lambda_{\text{opt}} < 1$ donc $\lambda = 1$ rentre dans la zone où le TEB varie peu.

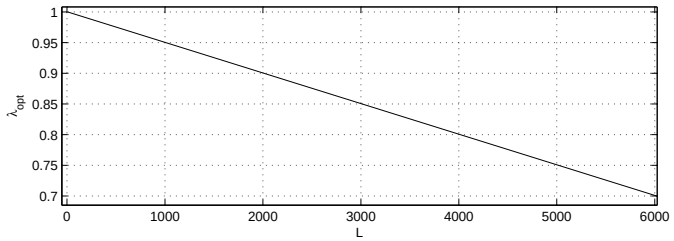

FIG. 1.9 – λ_{opt} en fonction de L, WNR=-3 dB, DWR=28 dB

1.3.2 Tatouage quantificatif

Les méthodes de tatouage par quantification sont un cas particulier de méthodes substitutives avec dictionnaire. On quantifie le signal hôte (ou une transformée de ce signal) selon un dictionnaire prédéfini de quantificateurs, correspondant à différents messages. Le décodage détermine la grille de quantification la plus proche de la composante de $\mathbf{z}$, et donc le message correspondant. Ce sont des méthodes avec "état de l'hôte connu" par opposition aux méthodes avec "statistique de l'hôte connue" (cf. paragraphe 1.5.5). Elles sont une implantation pratique du principe de tatouage informé, et plus précisément du principe de catégorisation aléatoire (cf. paragraphe 1.2.6). En effet, le dictionnaire de quantificateurs correspond à un dictionnaire de *bins* structuré et sous-optimal, de dimension raisonnable. Les techniques quantificatives les plus connues sont la Modulation d'Indices de Quantification (QIM) de Chen et Wornell et le Schéma de Costa Scalaire (SCS) de Eggers *et al.*

Principe du tatouage quantificatif

Quantification scalaire : on appelle quantification scalaire de pas Δ et d'indice de quantification $\tau < \Delta$ l'opération

$$Q_{\Delta,\tau}(x) = \Delta \text{Int}(\frac{x+\tau}{\Delta}) - \tau.$$

On appelle "grille de quantification" l'ensemble image de $Q_{\Delta,\tau}$: $\{l\Delta - \tau, l \in \mathbb{Z}\}$. Dans le tatouage quantificatif, x est une composante scalaire extraite du document, par exemple la luminance d'un pixel d'une image. Soit $q < \Delta$ les "détails" de x, reste de la division euclidienne de $x + \tau$ par Δ. On peut alors décomposer x en :

$$x = (l\Delta + \tau) + q - \tau = Q_{\Delta,\tau}(x) + q - \tau \tag{1.12}$$

Le tatouage quantificatif d'un bit d'information b sur le scalaire x consiste à substituer à x une valeur y située sur l'une de deux grilles $Q_{\Delta_{-1},\tau_{-1}}$ (correspondant à $b = -1$) ou Q_{Δ_1,τ_1} (correspondant à $b = 1$). Par exemple, on peut prendre $\Delta_{-1} = \Delta_1$, $\tau_{-1} = +\Delta/4$, $\tau_1 = -\Delta/4$. Cette substitution est représentée sur la *fig.* 1.10.

Le décodage consiste à déterminer la grille de quantification la plus proche du signal reçu z. Les deux grilles doivent donc être disjointes et le plus éloignées possible. Le choix de Q, τ et Δ permet d'optimiser l'éloignement des grilles et donc les performances à la détection. En l'absence d'attaque ($z = y$), z est situé exactement sur un point de la grille $\{l\Delta - \tau_b, l \in \mathbb{Z}\}$ et le décodage est parfait (TEB nul). On appelle cette propriété **rejet des interférences de l'hôte** : contrairement au cas du tatouage aveugle, le document hôte ne joue pas le rôle d'un bruit de transmission.

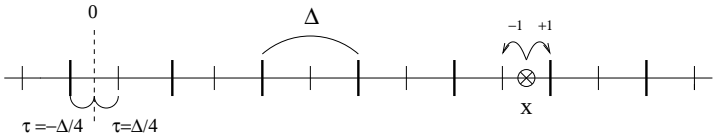

FIG. 1.10 – Quantification scalaire

Quantification vectorielle : l'opération de quantification vectorielle consiste à réduire un espace N-dimensionnel $\mathcal{E}$ décrivant le document en un espace $N - m$ dimensionnel $\mathcal{F}$. Par exemple, $\mathcal{E}$ peut représenter une image couleur dans l'espace RGB ($N = 3$) et $\mathcal{F}$ l'ensemble des 256 couleurs les plus utilisées ($N - m = 1$). La quantification vectorielle est donc habituellement une méthode de compression, ou de codage source. En pratique, $\mathcal{E}$ est découpé en sous-ensembles connexes appelées cellules de Voronoï. Le point de chaque cellule de Voronoï situé à distance égale de chaque frontière de la cellule est appelé "centroïde". $\mathcal{F}$ est l'ensemble des centroïdes. L'image de tout point d'une cellule par l'opération de quantification est donc son centroïde.

Le tatouage quantificatif vectoriel suit le même principe que dans le cas scalaire. Les techniques de quantification vectorielle utilisées en pratique sont souvent des grilles de quantification régulières, qui permettent d'atteindre la capacité du canal [ZSS02][ZD05][FTB04]. Par exemple, la grille hexagonale est optimale en deux dimensions. Dans le cas extrême (quantification à très haute dimension), on pourrait imaginer de quantifier l'image entière en un seul centroïde : on passerait de la dimension N à un scalaire. Il serait alors inutile d'avoir recours à une redondance. Lorsqu'on manipule un vecteur $\mathbf{x}$, on appelle également "quantification scalaire" le fait d'appliquer sur chaque dimension x_k du signal un quantificateur scalaire. On peut donc considérer cela comme une quantification vectorielle, mais dont le quantificateur est le produit cartésien des quantificateurs scalaires (cf. *fig.* 1.11).

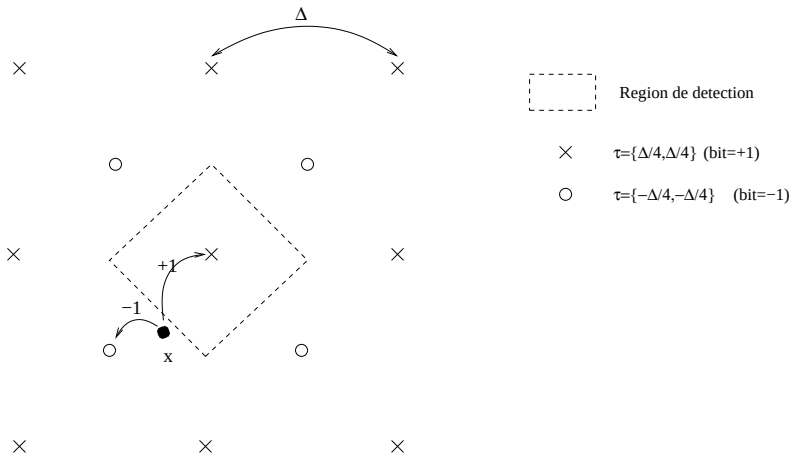

FIG. 1.11 – Quantification vectorielle avec pas uniforme

Modulation d'Indices de Quantification (QIM)

Le tatouage quantificatif a été introduit par Chen et Wornell sous le nom de **QIM** (*Quantization Index Modulation*) [CW01]. Cette appellation généralise explicitement la technique "historique" LSB (cf. paragraphe 1.2.7), qui peut être considérée comme une substitution d'un pixel d'une image par le résultat de deux quantificateurs grossiers : celui pour lequel le plan LSB (bit le moins important perceptuellement) vaut toujours 0, et celui pour lequel il vaut toujours 1. Dans la QIM, on étend ce principe à d'autres quantités scalaires x, d'autres quantificateurs Q et à la quantification vectorielle. Le choix des quantificateurs utilisés est déterminé, parmi un ensemble donné, par le message à insérer, d'où l'appellation "Modulation d'Indices de Quantification". Le pas de quantification Δ résulte d'un compromis entre les distorsions provoquées par l'insertion, la performance de détection et le nombre de mots contenus dans le dictionnaire. Les méthodes de la classe QIM sont optimales en termes de performance au décodage et de capacité d'insertion dans certains scénarios d'attaque comme l'AWGN.

Une implantation simple est proposée sous l'appellation *Dither Modulation* ou "modulation d'agitation" (**DM-QIM**). On y quantifie le signal sous la forme retenue dans (1.12) :

$$y_k = Q_{\Delta_k, b_k\tau_k}(x_k - b_k\tau_k) - b_k\tau_k$$

La particularité de la méthode est que le vecteur $\{\tau_k\}$ est ici un vecteur pseudo-aléatoire appelé *Dither-Vector* ou "bruit d'agitation". Par exemple, $\tau_k = \pm\Delta/4$ (au lieu de $\tau = \Delta/4$ constant dans (1.12)). On voit bien que **b** "module" $\{\tau_k\}$. L'implantation la plus simple est le DM-QIM avec quantification scalaire à pas uniforme ($\Delta_k = \Delta$).

Compensation des distorsions (DC-QIM)

La QIM peut être considérée comme une implantation sous-optimale du schéma de Costa (cf. paragraphe 1.2.6), avec $\mathbf{u} = Q_{\Delta,\tau}(\mathbf{x})$ représentant la version quantifiée du document et $\alpha = 1$. Une version optimale du schéma de Costa devrait utiliser un dictionnaire $\mathcal{U}$ aléatoire et choisir $\mathbf{u} \in \mathcal{U}$, mais la taille de $\mathcal{U}$ nécessaire rend son implantation impossible. La QIM est exactement un schéma de Costa sous-optimal lorsque les distorsions de l'attaque sont ignorées : WNR$\rightarrow \infty$. Chen et Wornell proposent donc une variante appelée Compensation des Distorsions (**DC-QIM**) qui prend

α en compte :

$$\mathbf{y} = Q_{\Delta/\alpha,\tau}(\mathbf{x}) + (1-\alpha)(\mathbf{x} - Q_{\Delta/\alpha,\tau}(\mathbf{x}))$$

$$\textit{i.e.} \quad \mathbf{y} = \mathbf{x} + \alpha(Q_{\Delta/\alpha,\tau}(\mathbf{x}) - \mathbf{x}), \quad \text{donc toujours } \mathbf{u} = \alpha Q_{\Delta/\alpha,\tau}(\mathbf{x})$$

Si en plus on utilise la modulation d'agitation, l'algorithme est appelé **DC-DM**. La modification de l'opérateur de quantification est représentée sur la *fig.* 1.12.

L'intérêt de la compensation des distorsions rejoint le principe de Costa : rendre maximale la robustesse au bruit, à distorsion donnée. En effet, on augmente le pas de quantification en transformant Δ en Δ/α. La distorsion augmente en $1/\alpha^2$, ainsi que la robustesse. Parallèlement, l'injection dans le tatouage de $(1-\alpha)(\mathbf{x} - Q_{\Delta/\alpha,\tau}(\mathbf{x}))$ diminue la distorsion. Soit d_1^2 la distance minimum entre deux grilles de quantification si $\alpha = 1$. Elle devient d_1^2/α^2 avec compensation des distorsions. Les interférences au décodage proviennent de l'attaque ($\sigma_{\mathbf{n}}^2$) et de la compensation des distorsions ($\sigma_{\mathbf{w}}^2/\alpha^2$). On retrouve le α_{opt} de Costa en maximisant par rapport à α le rapport suivant :

$$\frac{d_1^2/\alpha^2}{(1-\alpha)^2 \frac{\sigma_{\mathbf{w}}^2}{\alpha^2} + \sigma_{\mathbf{n}}^2}$$

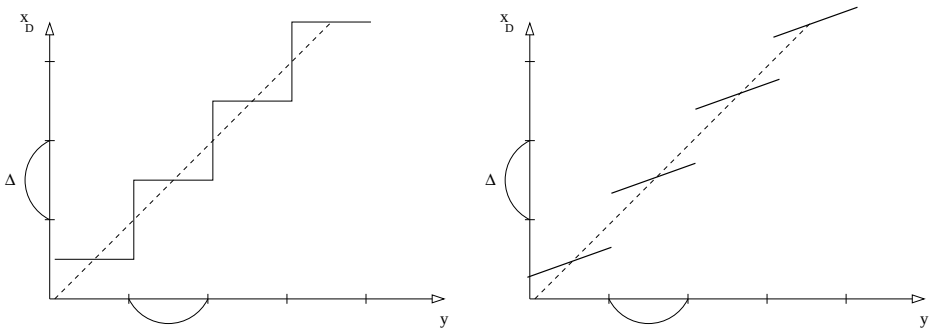

FIG. 1.12 – Entrée et sortie du quantificateur, sans ou avec Compensation des distorsions

Schéma de Costa Scalaire (SCS)

Eggers *et al.* [EBTG03], ayant eux aussi remarqué le lien entre QIM et schéma de Costa, ont proposé sous le nom de Schéma de Costa Scalaire **SCS** (*Scalar Costa Scheme*) une implantation sous-optimale explicite du schéma de Costa. Bien qu'il soit très proche d'une DC-DM avec une quantification scalaire à pas uniforme, le SCS est très populaire par sa présentation plus simple. Le SCS est fondé sur la grille suivante :

$$\mathcal{M}_{m,k} = \left(\frac{\Delta}{\alpha}\mathbb{Z} - m\frac{\Delta/\alpha}{|\mathcal{M}|} - \tau_k \right)$$

où $m \in \{1 \dots |\mathcal{M}|\}$ est l'indice du symbole à transmettre et $|\mathcal{M}|$ la taille du dictionnaire. Soit $Q_{\mathcal{M}_m}$ l'opérateur de quantification dont l'espace image est le sous-dictionnaire $\mathcal{M}_m$. Le principe est ici d'ajouter au mot de code $U = \alpha Q_{\mathcal{M}_{m,k}}(x_k)$ une fraction $(1-\alpha)$ de l'erreur de quantification $x_k - Q_{\mathcal{M}_{m,k}}(x_k)$. On remarque qu'en pratique U dépend de α, ce qui n'est pas précisé dans le schéma de Costa, mais qui est dicté par l'implantation. Le tatouage est donc :

$$\mathbf{x} \to \mathbf{y} = U + (1-\alpha)\mathbf{x} \quad \text{et} \quad w_k = \alpha(Q_{\mathcal{M}_m}(w_k) - w_k + m\frac{\Delta}{|\mathcal{M}|} + \tau_k)$$

ce qui correspond à $Q_{\mathcal{M}_m}(x_k) = Q_{\Delta/\alpha}(x_k - m\frac{\Delta\alpha}{|\mathcal{M}|} - \tau_k)$. Il s'agit bien de DC-DM, avec un vecteur d'agitation égal à $m\frac{\Delta\alpha}{|\mathcal{M}|} - \tau_k$. La sécurité de l'algorithme vient de l'utilisation des $\tau_k \in [0, 1)$ (cf. paragraphe 1.4). La transmission d'un signal $|\mathcal{M}|$-aire ne se justifie que si WNR$\geq$ 4 dB (bruit faible), ce qui correspond à un tatouage à haut débit. En pratique, on utilisera donc toujours un SCS binaire : $|\mathcal{M}| = 2$ et m correspond directement à b_k. Au décodage, on retrouve $\mathcal{M}_{\hat{b}_k,k}$ par :

$$\hat{d}_k = Q_{\mathcal{M}_m}(y_k) - (b_k\frac{\Delta}{|\mathcal{M}|} + \tau_k)$$

donc dans le cas binaire $\hat{d}_k \leq \frac{\Delta}{2}$, et

$$\hat{b}_k = \text{argmin}_{b_k=0,1}(|\hat{d}_k - \frac{b_k}{2}|)$$

Ce schéma suppose que les attaques sont exclusivement basées sur l'ajout de bruit gaussien, et que l'on connaît $\sigma^2_{\mathbf{n}}$ à l'insertion. Dans l'étude théorique des performances, on suppose le document hôte uniformément distribué sur les centroïdes, ce qui est faux si Δ est grand (car alors on aurait $\sigma^2_{\mathbf{x}} = \infty$). Avec cette hypothèse, l'erreur de quantification est $\frac{\Delta}{12}$ et $\sigma^2_{\mathbf{w}} = \frac{\Delta\alpha}{12}$. Comme le schéma est sous-optimal, $\alpha_{\text{SCS}} \neq \alpha_{\text{opt}}$. En calculant numériquement les intégrales présentes dans le calcul théorique des performances, on obtient

$$\alpha_{\text{SCS}} = \sqrt{\frac{\sigma^2_{\mathbf{w}}}{\sigma^2_{\mathbf{w}} + 2.71\sigma^2_{\mathbf{n}}}} \ .$$

Pour l'imperceptibilité, il est important de soustraire le signal d'agitation après quantification. Ainsi, on quantifie toujours au centroïde le plus proche (cf. *fig.* 1.13). C'est le principe du signal d'agitation soustractif utilisé en quantification pour diminuer la distorsion [WLVW00].

Des signaux d'agitation non soustractifs sont également utilisés en quantification [WLVW00]. Ils permettent notamment d'imposer des propriétés statistiques à l'erreur de quantification. Cependant, ils ne garantissent pas l'indépendance entre le signal quantifié et l'erreur de quantification, à moins d'augmenter la distorsion. Ceci pourrait nuire à la sécurité d'une technique de tatouage. Seuls les travaux de Eggers et Girod font mention de l'utilisation d'un signal d'agitation non soustractif pour le tatouage [EG00]. Ils proposent que le tatouage serve de signal d'agitation, par exemple conjointement à une compression JPEG par quantification non soustractive. Il ne s'agit pas d'une technique de catégorisation aléatoire.

Lien entre ISS et quantification scalaire

En présence de fort bruit, le nombre de centroïdes nécessaires dépend de $\sigma_{\mathbf{x}}/\Delta$. Or $\alpha_{\text{opt}} \to 0$ pour un WNR faible. Pour garder un DWR constant à faible WNR, on doit donc augmenter Δ afin que $\alpha(Q_{\mathcal{M}_m}(\mathbf{x}) - \mathbf{x})$ soit encore grand. Le nombre de centroïdes diminue donc avec $1/\Delta$ et il suffit d'un mot de code par symbole. Selon ce principe, le SCS binaire se réduit à 2 centroïdes $\{-\frac{\Delta}{4}, \frac{\Delta}{4}\}$ et l'on a, avec α petit et Δ grand [PFPG05] :

$$\mathbf{w} = \alpha\left((-1)^b\frac{\Delta}{4} - \mathbf{x}\right) \text{ avec } m = \pm 1 \ . \tag{1.13}$$

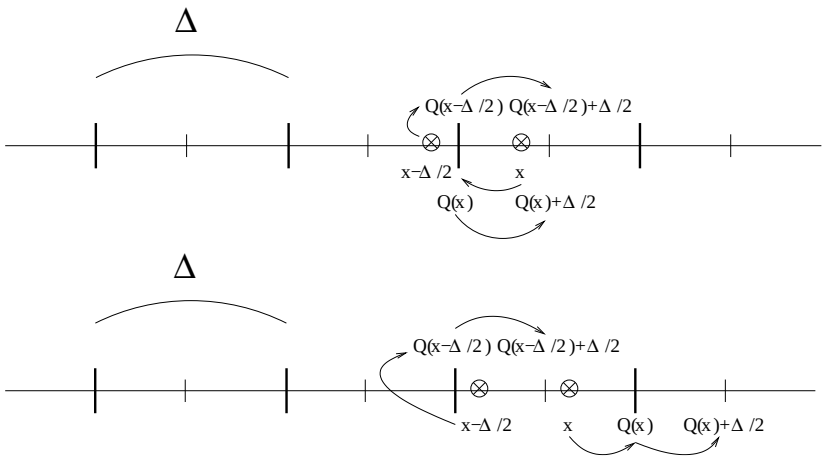

FIG. 1.13 – Haut : cas favorable ; Bas : cas défavorable

L'ISS est relié par Pérez-Freire *et al.* dans [PFPG05] au *Distortion Compensated-Spread Spectrum* (DC-SS), une technique issue du principe de quantification scalaire. Ces deux méthodes permettent de faire un pont entre les principes d'étalement de spectre et de quantification. Le DC-SS reprend le schéma de l'équation (1.13), dérivé du SCS, en tant que schéma de tatouage, même si le WNR n'est pas faible. Comme il n'y a que deux centroïdes, on peut ramener ce schéma à de l'étalement de spectre généralisé. Pour compléter l'analogie, il faudrait rajouter une redondance P et une clé secrète, constituée par le paramètre τ_k ici omis. On obtient les mêmes calculs de performance et le même choix de paramètres optimaux que dans le cas de LISS avec α fixé. Dans les deux cas, le débit accessible du schéma est

$$R(\alpha,\lambda) \simeq \frac{1}{2}\log_2(1+\text{SNR}), \text{ où SNR} \triangleq \frac{\sigma_{\mathbf{c}}^2 - \lambda\sigma_{\mathbf{x}}^2}{(1-\lambda)^2\sigma_{\mathbf{x}}^2 + \sigma_{\mathbf{n}}^2}.$$

Dans les deux cas, l'approximation n'est valable que pour WNR faible. Si le SCS et l'ISS ont des performances identiques à faibles WNR, dans le cas d'un bruit faible le nombre de centroïdes augmente dans le SCS, d'où des performances supérieures.

Tatouage à faible débit : redondance et étalement (STDM, ST-SCS et QP)

Si en l'absence de bruit l'ensemble des méthodes quantificatives présentent un TEB nul et une capacité infinie, dès que le document est soumis à des attaques, les performances diminuent fortement. On doit donc introduire une redondance, par exemple sous la forme d'une insertion répétée pour un même bit d'information (**redondance par répétition**). Une deuxième solution consiste à appliquer une technique de codage canal, avec les mêmes limitations que pour l'étalement de spectre.

Eggers *et al.* montrent [EBTG03] que l'ajout de **redondance par étalement** offre de bien meilleures performances. En effet, la composante du bruit qui est orthogonale au code ne perturbe pas le décodage par corrélation (cf. *fig.* 1.14). On a alors, à TEB égal, $\text{WNR}_P = P\text{WNR}_1$, où P est le facteur d'étalement. Le défaut de l'étalement par rapport à la répétition est le problème de la synchronisation nécessaire entre $\mathbf{z}$ et $\mathbf{c}$. De plus, l'approximation de la distribution du support par une distribution uniforme tient encore moins dans le domaine projeté. L'algorithme proposé par Eggers s'appelle **ST-SCS** (*Spread Transform SCS*). L'idée de redondance par étalement était également

présente dans les travaux de Chen et Wornell sous le nom de **STDM** ou *Spread Transform Dither Modulation* [CW01]. DC-STDM est similaire à ST-SCS.

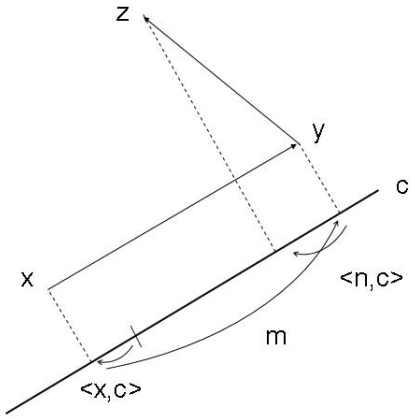

FIG. 1.14 – Impact de la redondance par étalement sur le bruit

Une variante simple du STDM est proposée par Pérez-Gonzalez et Balado [PGB02] sous le nom de Projection Quantifiée (*Quantized Projection*, **QP**) et consiste à quantifier

$$x =< \underline{x}_{P_s}, \mathbf{c} >= Q_{\Delta,\tau}(x) + q - \tau \quad \longrightarrow \quad y = Q_{\Delta,\tau}(x)$$

où $\underline{x}_{P_s}$ est un vecteur de P_s pixels de l'image et **c** une clé formée d'un pseudo-bruit de longueur P_s. Ce schéma correspond donc à

$$\underline{y}_{P_s} = \underline{x}_{P_s} - (q - \tau)\mathbf{c} \ ,$$

c'est-à-dire à l'ajout d'un motif pseudo-aléatoire. La différence avec l'étalement de spectre réside dans la faible taille de **c** (par exemple $P_s = 20$), donc le produit de corrélation $z =< \underline{z}_P, \mathbf{c} >\neq 0$ pour **c** quelconque. La réception consiste à effectuer un test d'hypothèse pour déterminer si z est un point de la grille $Q_{\Delta,\tau}$. La différence entre STDM et QP réside donc dans la formulation de la fonction de reconstruction (*unprojection*), même si dans le cas de base les deux méthodes sont identiques.

Les trois techniques précédentes consistent à quantifier un scalaire (dimension 1) au lieu d'un vecteur de dimension P. La technique de projection sur un sous-espace **SSP** (*Subspace Projection*) généralise cette approche en quantifiant un vecteur de dimension $N_P \leq N$ [FTB04]. Ce vecteur est calculé en projetant **x** par multiplication par une matrice orthonormale de dimension $N \times N_P$. On peut ensuite appliquer une technique de quantification vectorielle, plus performante que la quantification scalaire de STDM [FTB04].

Accès multiple : pour insérer plusieurs tatouages comme dans le cas de l'étalement de spectre, on a deux solutions. La première est d'insérer les messages sur des supports disjoints, par exemple avec un SCS simple (division temporelle ou spatiale). La seconde est d'utiliser une variante avec étalement où les vecteurs d'étalement sont orthogonaux (division par code). Le problème de l'accès multiple pour le schéma de Costa est étudié dans [ZP06] et des codes pratiques permettant à la fois le rejet des

interférences de l'hôte et l'accès multiple sont proposés dans [ZPD05].

Tatouage et compression conjointe : un schéma de tatouage et compression jointe est présenté dans [GM02][GM04]. En effet, certains algorithmes de compression quantifient les coefficients de sous-bandes de la décomposition en ondelettes d'une image. Le pas de quantification dépend d'un débit adapté à la distorsion pour chaque coefficient. Le principe du tatouage joint est de quantifier selon des sous-dictionnaires dépendant des mots de code, comme dans la QIM. Ici, les auteurs adaptent un algorithme de compression utilisant la quantification vectorielle par grille régulière (LVQ). Ils montrent que la modulation DM introduit une augmentation de l'entropie du signal, ce qui nuit à la compression, et proposent une nouvelle technique de modulation (*Vector Dead Zone*).

Performances au décodage et à la détection

Distorsion : dans le cadre du tatouage informé, le tatouage est dépendant de l'hôte, on ne peut donc pas estimer *a priori* la puissance du tatouage et le DWR. On a donc *a posteriori* la distorsion suivante, avec la puissance du tatouage D_w mesurée au sens des moindres carrés (MSE) :

$$\text{DWR} = 10\log_{10}\frac{\sigma_{\mathbf{x}}^2}{D_w}, \text{ avec :}$$

$$D_w = \mathrm{E}[\|\mathbf{y}-\mathbf{x}\|^2] = \frac{1}{\mathcal{M}}\sum_{m\in\mathcal{M}}\int (y-x)^2 f_{\mathbf{x}}(x)\mathrm{d}x \tag{1.14}$$

Décodage : le schéma de Costa est conçu pour optimiser la capacité de l'algorithme de tatouage dans le cas d'un hôte gaussien et d'une attaque gaussienne. Le débit accessible tend vers la capacité infinie lorsque $N \to +\infty$. Si $\sigma_{\mathbf{x}} < \sigma_{\mathbf{w}}$, les méthodes quantificatives sont donc beaucoup plus robustes à l'attaque AWGN que l'étalement de spectre [PGBM03]. Une idée reçue était qu'en théorie les méthodes quantificatives sont moins robustes que l'étalement de spectre lorsque WNR est faible (forte attaque). Il s'agit d'une erreur due au fait que la densité de probabilité du signal hôte à l'intérieur de chaque cellule de quantification est supposée uniforme [PFPG05][PFPGV06]. Les centroïdes sont alors équiprobables, ce qui impliquerait que le document a une variance infinie. De plus, la distribution uniforme n'est pas adaptée lorsque le rapport $\sigma_{\mathbf{x}}^2/V$ est grand, où V est le volume d'une cellule de quantification, lié à Δ. C'est par exemple le cas lorsqu'une transformée d'étalement est utilisée. En tenant compte de la connaissance du document hôte à l'insertion dans les calculs, les performances du SCS face à l'attaque AWGN ne sont jamais inférieures à celles de l'étalement de spectre.

Détection : heuristiquement, le problème de la détection est souvent résolu en étudiant les performances à la détection du schéma optimal (ou sous-optimal en pratique) construit pour le décodage. Ces performances, qui elles ne sont pas forcément optimales, sont calculées notamment pour QP [PFCPG05a] et DC-QIM [BDBT06b]. L'optimisation de la détection pour le DC-QIM est étudiée dans [BDBT06b]. Notamment, on recalcule le paramètre α du DC-QIM qui maximise la performance de la courbe COR, et qui donc peut être différent de celui maximisant la performance du décodage. Les courbes COR des schémas quantificatifs ne sont pas symétriques par rapport à la

diagonale, contrairement au DS. Le résultat de l'étude est que α_{opt} est identique pour le problème de la détection et du décodage. Cependant, on n'a pas montré l'optimalité pour la détection de DC-QIM en tant que schéma. On peut construire des quantificateurs optimaux pour le problème de la détection, qui fournissent de meilleures performances que ceux optimisant le décodage [BDBT06a].

Scénarios d'attaque : l'étude de la robustesse aux attaques répond à deux scénarios. Lorsque l'attaque est connue, une compensation des distorsions est possible. C'est ce qui est effectué dans DC-DM et SCS avec l'attaque AWGN si $\sigma^2_{\mathbf{n}}$ est connu. Pour d'autres attaques, une solution est par exemple d'estimer le $\sigma^2_{\mathbf{n}}$ équivalent. Lorsque l'attaque ou ses paramètres sont inconnus, il est impossible d'effectuer une compensation des distorsions. En tatouage, on suppose cependant que dans le "jeu", l'intérêt du pirate est toujours d'utiliser la puissance d'attaque maximale qui lui est permise, ce qui permet de choisir α. Une solution consiste alors à modéliser l'attaque par un AWGN, dont on fixe la variance $\sigma^2_{\mathbf{n}}$ maximale. Bien que les méthodes à compensations des distorsions améliorent les performances, ce n'est plus vrai dans le domaine projeté pour un faible WNR [PGBM03]. Dans ce cas, on peut fixer $\alpha = 1$ dans ST-SCS sans pénaliser le décodage. De manière générale, dans un scénario pratique, *i.e.* WNR proche de 1, pour P suffisamment grand, α_{opt} est proche de 1. En effet, pour ST-SCS $\alpha_{\text{opt}} = \sqrt{\frac{P\sigma^2_{\mathbf{w}}}{P\sigma^2_{\mathbf{w}}+\sigma^2_{\mathbf{n}}}} = \sqrt{\frac{1}{1+\frac{2.71}{P\text{WNR}}}}$. Donc si PWNR est grand devant 2.71, α_{opt} est proche de 1. Par exemple, pour WNR=-3 dB, $P = 50$, $\alpha_{\text{opt}} = 0.95$ (cf. *fig.* 1.15).

STDM est très robuste à l'attaque de quantification, donc à la compression JPEG [BBP04]. L'attaque de rognage aléatoire de pixels a plus d'impact sur ST-SCS que sur DS [CPG06].

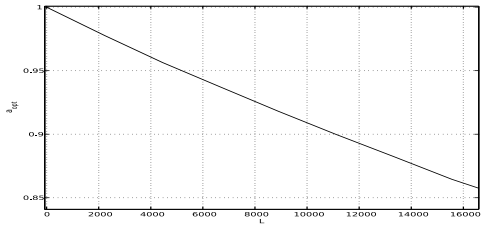

FIG. 1.15 – α_{opt} en fonction de L, WNR=-3 dB

Robustesse aux attaques valumétriques et contre-mesures

Les techniques quantificatives sont particulièrement vulnérables aux transformation valumétriques. Par exemple, l'attaque de gain ([BBP04] pour STDM, [PFCPG05a] pour QP) consiste à effectuer $\mathbf{z} = \rho(\mathbf{y} + \mathbf{n})$. Avec un simple facteur d'échelle ($\mathbf{n} = 0$), le pas de quantification Δ devient $\Delta' = \rho\Delta$ et le décodage avec Q_Δ est impossible. Un facteur $\rho = 1.1$ conduit déjà à TEB=0.5.

Devant la fréquence de ces attaques, des contre-mesures ont été imaginées. La première solution consiste à estimer la transformation valumétrique avant détection afin de calculer Δ'. L'estimation repose sur l'analyse de la transformée de Fourier de l'histogramme (technique également applicable à la transformation gamma non linéaire) [EBG02] et est éventuellement formulable comme un estimateur du maximum de vraisemblance [LS04]. Une autre solution appelée "insertion proportionnelle

à l'hôte" consiste à quantifier une composante qui sera affectée de manière linéaire, ou logarithmique, par la transformation. Pérez-Gonzalez *et al.* [PGMBA05] proposent la technique de Rational Dither Modulation (**RDM**), efficace pour des transformations valumétriques affines. Soit $\underline{y}_k = [y_{k-N_v}, \ldots, y_{k-1}]$ les N_v points précédant y_k (approche causale). Soit $g : \mathbb{R}^{N_v} \longrightarrow \mathbb{R}$ une fonction vérifiant la propriété d'homogénéité :

$$g(\rho \underline{y}_k) = \rho g(\underline{y}_k)$$

On insère alors

$$y_k = g(\underline{y}_k)(Q_{\Delta, b_k\tau_k}\left(\frac{x_k}{g(\underline{y}_k)}\right) - b_k\tau_k)$$

Au décodage, on calcule

$$\hat{b_k} = \arg\min_{-1,1} |\frac{z}{g(\underline{z}_k)} - Q_{\Delta, b_k\tau_k}\left(\frac{z_k}{g(\underline{z}_k)}\right) + b_k\tau_k|^2$$

Donc en cas de simple multiplication par un facteur d'échelle,

$$\hat{b_k} = \arg\min_{-1,1} |\frac{\rho y}{\rho g(\underline{y}_k)} - Q_{\Delta, b_k\tau_k}\left(\frac{\rho y_k}{\rho g(\underline{y}_k)}\right) + b_k\tau_k|^2 = b_k$$

g n'est pas nécessairement linéaire. On peut par exemple utiliser la norme $\mathcal{L}_\beta$ définie par :

$$g(\underline{y}_k) = \left(\frac{1}{N_v}\sum_{j=k-M}^{k-1} |y_j|^\beta\right)^{1/\beta}$$

Si $\beta = 1$, l'insertion est proportionnelle à la moyenne de $\underline{y}_k$, si $\beta = 2$ elle est proportionnelle à son écart-type. Le pas Δ de départ est adapté pour satisfaire le DWR visé :

$$\Delta = 2\sqrt{3\left(1 - \frac{\sigma_{\mathbf{x}}^2}{\sigma_{\mathbf{y}}^2}\right)}$$

Les performances tendent vers celles de QIM lorsque $N_v \to +\infty$ et sont supérieures à celles de LISS. Dans [PGMBA05], RDM est même généralisée en une classe de techniques QIM invariantes à l'attaque de gain, appelée GI-QIM (*Gain-Invariant QIM*) : RDM $|\mathcal{M}|$-aire, par blocs, à transformation d'étalement, compensation des distorsions...

La technique de "quantification adaptative" consiste à adapter Δ au signal pour des raisons de robustesse et d'imperceptibilité [OKS04]. On a alors localement sur le bloc $[x_1, \ldots, x_{N_v}]$:

$$\Delta_x = \frac{\Delta}{N_v}\sum_{k=1}^{N_v} x_k$$

Le pas est estimé à la détection :

$$\Delta_y = \Delta_x + \frac{\Delta}{N_v}\sum_{k=1}^{N_v}(w_k + b_k)$$

Cette technique est très proche de RDM, avec g qui est une moyenne locale et $\underline{y}_k$ qui n'est pas causale. Les techniques précédentes se limitent aux transformations affines. P. Bas [Bas05] propose de quantifier la valeur médiane de triplets de points (il y a donc une perte de capacité), pour résister aux transformations non affines. Le défaut commun à ces approches est leur moins grande robustesse à d'autres attaques comme l'AWGN.

1.4 Principes de la sécurité d'un algorithme de tatouage

On a différencié les attaques sur la robustesse, destinées à augmenter la probabilité d'erreur, et les attaques sur la sécurité, qui concernent la connaissance des clés secrètes (cf. paragraphe 1.2.3). L'étude de la sécurité d'un algorithme de tatouage a connu de grandes avancées durant les deux dernières années. Dans cette partie, nous présenterons les attaques classiques sur la sécurité, puis les principes théoriques qui permettent d'établir le niveau de sécurité d'un algorithme, ainsi que des attaques pratiques inspirées de cette étude. Ce cadre théorique sera ensuite appliqué aux techniques de tatouage utilisant les filtres LPTV et l'interpolation dans les parties 2.6 et 4.7.

1.4.1 Attaques classiques sur la sécurité

Historiquement, on a identifié les attaques cryptographiques, celles liés au protocole d'application dont l'algorithme de tatouage fait partie, les attaques de collusion entre utilisateurs et l'attaque de sensibilité.

Les attaques cryptographiques interviennent souvent dans la transmission de la clé entre l'encodeur et le décodeur. Celle-ci doit se faire par des protocoles cryptographiques conventionnels, et est donc vulnérable aux attaques cryptographiques traditionnelles (*man-in-the-middle*...). Afin d'éviter un possible décryptage de la clé d'encodage, le **tatouage asymétrique** repose sur l'utilisation de deux clés : une clé $\mathbf{k}_I$ privée pour l'insertion et une clé $\mathbf{k}_D$ publique pour la détection. $\mathbf{k}_D$ est issue de $\mathbf{k}_I$ par une transformation non inversible. N'importe quel utilisateur peut détecter le tatouage en connaissant $\mathbf{k}_D$, mais seule la connaissance de $\mathbf{k}_I$ permet d'enlever ou modifier $\mathbf{w}$. Par exemple, Furon et Duhamel [FD03] proposent d'insérer un signal $\mathbf{w}$ coloré spectralement puis étalé par entrelaceur. Seule la densité spectrale de puissance de ce signal peut être détectée, le signal lui-même ne pouvant être retrouvé. Le signal complet est la clé privée. Dans le même esprit, L. de C.T. Gomes [dCTG02] propose d'insérer un signal $\mathbf{w}$ cyclostationnaire entrelacé. Après entrelacement inverse, seule la cyclofréquence peut être détectée. Ces deux exemples concernent le tatouage d'un document sonore. Les fonctions de détection de tous les schémas asymétriques peuvent s'exprimer sous une forme quadratique, et sont tous moins robustes que les schémas symétriques [Fur02]. Le **tatouage à clé publique** a pour but d'utiliser une clé publique permettant d'utiliser un décodeur classique, mais qui ne contienne pas assez d'information pour enlever le tatouage. Cependant, aucun algorithme pratique de tatouage à clé publique n'a pu être construit. Enfin, le **tatouage à connaissance nulle** consiste à convaincre un usager qu'un tatouage est présent sans rien lui révéler du tatouage lui-même, par un protocole de cryptographie.

Un exemple intéressant d'**attaque de protocole** est celle imaginée par S. Craver [Cra98] pour l'application de preuve de propriété. On y envisage la situation où le pirate revendique la possession d'un copyright sur un document tatoué $\mathbf{y}_A = \mathbf{x}_A + \mathbf{w}_A$ en exhibant son propre tatouage $\mathbf{w}_B$ obtenu par soustraction à partir du document tatoué : $\mathbf{y}_B = \mathbf{x}_B + \mathbf{w}_B$ avec $\mathbf{x}_B = \mathbf{x}_A + \mathbf{w}_A - \mathbf{w}_B$, le pirate déclarant que $\mathbf{x}_B$ est le document original. Le document contiendra donc à la fois le tatouage du propriétaire de l'œuvre et celui du pirate. Ceci introduit la nécessité d'algorithme de tatouage non-inversible, où l'étape de soustraction est impossible. Une solution applicable à n'importe quel algorithme est que $\mathbf{w}$ dépende à la fois de $\mathbf{m}$ et de $\mathbf{x}$. Ainsi $\mathbf{w}_B$ dépendra de $\mathbf{w}_A$.

Les attaques dites de moyennage ou de **collusion**, s'appliquent au cas où l'on pos-

sède de nombreuses versions d'un même document tatouées avec différentes clés : c'est le cas des vidéos [DD05a], ou de l'application d'estampillage. L'attaque consiste à faire une moyenne de ces réalisations, ou à former un nouveau document à partir de fragments des différentes réalisations (attaque aussi appelée "mosaïque") [KP03a]. Une variante consiste à remplacer chaque bloc du signal par un autre bloc ou une combinaison d'autres blocs similaires perceptuellement [DD05b]. Là encore, une solution consiste à utiliser un tatouage dépendant du document hôte. Dans [BS95], on propose d'utiliser des codes sûrs à la collusion : la somme de plusieurs codes contient une information sur chacun des codes la composant. Par exemple, $\mathbf{w}$ doit posséder les mêmes propriétés d'auto-similarité que $\mathbf{x}$ pour résister à l'attaque de remplacement de blocs. La collusion peut également conduire à une attaque sur la sécurité : le moyennage sert à estimer le tatouage. Si une information adjacente est utilisée à l'encodage, cette attaque échoue.

L'**attaque de sensibilité** suppose que le pirate possède l'algorithme de détection, mais pas la clé. Il peut alors soumettre divers documents tatoués au détecteur et observer la réponse. Le pirate peut ainsi déterminer la frontière de détection de $\mathbf{w}$, et choisir pour $\mathbf{z}$ le document le plus proche de $\mathbf{y}$ qui soit hors de la zone de détection. Une contremesure possible consiste à utiliser une frontière de détection fractale [MT02][BBF03]. Ainsi, il est impossible de trouver une courbe paramétrable à partir des tangentes locales à la réponse du détecteur. L'attaque de sensibilité est généralisée et étudiée de façon théorique dans [CPFPG06].

1.4.2 Etude théorique du niveau de sécurité

Les attaques concernées par cette étude sont celles visant à estimer $\mathbf{k}$, dans le but d'enlever ou de modifier le tatouage.

Formalisation des attaques sur la sécurité

La formalisation du problème de la sécurité dans le domaine du tatouage est récente, les chercheurs s'étant plutôt concentrés jusqu'ici sur le problème de la robustesse. La sécurité n'était abordée initialement que sous un aspect cryptographique. Les techniques d'étalement de spectre étaient supposées sûres, sans que les spécificités du tatouage numérique soient prises en compte. Pourtant, Mihcak *et al.* ont montré que 90% de la clé secrète d'un tatouage DS peuvent être révélés par une attaque d'estimation (cf. partie 1.2.3) en prenant en compte les modèles statistiques du tatouage et de l'hôte, et utilisant un filtrage de Wiener [MVK02]. Barni *et al.* ont fourni en 2003 un cadre général pour l'étude du problème de la sécurité, utilisant notamment la théorie des jeux [BBF03]. Voloshynovskiy *et al.* [TVK$^+$05] s'intéressent aux algorithmes résistants aux attaques géométriques, tels que l'insertion d'une mire (cf. paragraphe 1.5.4). Le niveau de sécurité utilisé est la borne supérieure du cardinal de l'espace de recherche pour l'attaquant, dans les différents scénarios. La formalisation du problème de la sécurité d'un schéma de tatouage a connu une grande avancée avec les travaux de Cayre, Furon et Fontaine parus fin 2005 [CFF05], simultanément aux efforts de Comesaña, Pérez-Freire et Pérez-Gonzalez [CPFPG05b][CPFPG05a][PFCPG05b].

La caractéristique essentielle des attaques sur la sécurité est que le pirate a accès à N_o documents $\{\mathbf{y}^1, \ldots, \mathbf{y}^{N_o}\}$ tatoués avec la même clé $\mathbf{k}$. On dit qu'un algorithme est à **couverture parfaite** (*perfect covering*) si aucune information ne fuit de $\{\mathbf{y}^1, \ldots, \mathbf{y}^{N_o}\}$ au sujet de $\mathbf{k}$. Trois attaques sont identifiées [CFF05]. Dans l'**Attaque Connaissant le Message (KMA)** ou *Known Message Attack*, le pirate a

accès à $\{\mathbf{y}^1, \ldots, \mathbf{y}^{N_o}\}$ et aux N_o messages correspondants $\{M^1, \ldots, M^{N_o}\}$. Dans l'**Attaque Connaissant l'Original (KOA)** ou *Known Original Attack*, le pirate a accès à $\{\mathbf{y}^1, \ldots, \mathbf{y}^{N_o}\}$ et aux originaux $\{\mathbf{x}^1, \ldots, \mathbf{x}^{N_o}\}$, donc aux tatouages $\{\mathbf{w}^1, \ldots, \mathbf{w}^{N_o}\}$. Enfin, dans l'**Attaque connaissant Seulement le document Tatoué (WOA)** ou *Watermarked Only Attack*, le pirate n'a accès qu'à $\{\mathbf{y}^1, \ldots, \mathbf{y}^{N_o}\}$. Dans [CPFPG05a], KMA et KOA sont vus comme des cas particuliers de l'Attaque avec Estimation de l'Original (EOA), avec des erreurs différentes sur l'estimation de l'hôte.

Deux approches sont en concurrence : celle utilisant la théorie de l'information et de la cryptographie de Shannon, prônée par Pérez-Gonzalez *et al.*, et celle utilisant l'information de Fisher, préférée par Furon *et al.*

Approche de Shannon

Le critère choisi dans [CPFPG05b][CPFPG05a][PFCPG05b] pour mesurer la sécurité est l'information mutuelle $\mathrm{I}(\mathbf{y}^1, \ldots, \mathbf{y}^{N_o}; K)$ entre $\mathbf{k}$ et les N_o observations de différents documents tatoués avec $\mathbf{k}$. L'équivoque (ou "ambiguïté", pour *equivocation*), est l'incertitude restante sur $\mathbf{k}$ après N_o observations :

$$\mathrm{H}(\mathbf{k}|\mathbf{y}^1, \ldots, \mathbf{y}^{N_o}) = \mathrm{H}(\mathbf{k}) - \mathrm{I}(\mathbf{y}^1, \ldots, \mathbf{y}^{N_o}; \mathbf{k}) \ .$$

Ces deux critères sont issus des travaux de Shannon sur la cryptographie. Pour refléter à la fois l'incertitude *a priori* sur la clé et *a posteriori* après observations, on doit connaître deux des trois quantités. Shannon a également défini la distance d'unicité N_o^*, qui est la première valeur de N_o pour laquelle l'équivoque devient nulle. Elle mesure donc le niveau de sécurité : c'est une borne N_o^* du nombre minimum d'observations nécessaires à l'attaquant pour estimer la clé. La couverture parfaite correspond à $\mathrm{I}(\mathbf{y}^1, \ldots, \mathbf{y}^{N_o}; \mathbf{k}) = 0$.

Approche de Fisher

Cayre, Fontaine et Furon utilisent les matrices d'information de Fisher de préférence à l'équivoque dans le cas d'un secret $\mathbf{k}$ continu, ce qui est souvent le cas en tatouage, notamment lorsque le code d'étalement des techniques d'étalement de spectre suit une loi normale. En effet, l'entropie n'a pas de sens physique pour des variables continues [CFF05] : elle peut être positive, négative, ou ne pas exister ; elle n'est définie que par analogie avec le cas discret. Il suffit cependant de remplacer l'entropie par l'entropie différentielle dans l'approche de Shannon pour lever cette réserve [PFCTPPG06].

Soit O l'ensemble des données observées par l'attaquant. La Matrice d'Information de Fisher (FIM) est :

$$\mathrm{FIM}(\mathbf{k}) \triangleq \mathrm{E}[(\nabla_{\mathbf{k}} \log \mathrm{p}_O(o; \mathbf{k}))(\nabla_{\mathbf{k}} \log \mathrm{p}_O(o; \mathbf{k}))^T]$$

où $\nabla_{\mathbf{k}}$ est le gradient de $\mathbf{k}$: $\nabla_{\mathbf{k}}(\partial/\partial k_1, \ldots, \partial/\partial k_{N_K})$. Si la FIM est inversible, la borne de Cramér-Rao, borne inférieure de la matrice de covariance $\mathcal{R}_{\hat{\mathbf{k}}}$ de tout estimateur non biaisé de $\mathbf{k}$ à partir de O est la suivante :

$$\mathcal{R}_{\hat{\mathbf{k}}} \geq \mathrm{FIM}(\mathbf{k})^{-1}$$

Or l'erreur d'estimation $\mathrm{E}[\|\hat{\mathbf{k}} - \mathbf{k}\|^2]$ est la trace de $\mathcal{R}_{\hat{\mathbf{k}}}$. De plus, $\mathrm{tr}(\mathrm{FIM}(\mathbf{k})^{-1})$ décroît en $1/N_o$ et on veut calculer un niveau de sécurité indépendant de N_o. On définit donc :

$$N_o^* = N_o \mathrm{tr}\ (\mathrm{FIM}(\mathbf{k})^{-1})$$

Il s'agit d'une borne sur la difficulté d'estimation de la clé, même si elle n'est pas similaire à celle définie par Shannon. On peut juste dire que le nombre d'observations dont l'attaquant a besoin pour estimer la clé est en $O(N_o^*)$.

On trouve dans [CPFPG05a] la relation entre l'approche de Shannon et celle de Fisher. L'équivoque peut notamment être considérée comme l'entropie d'une variable gaussienne de matrice de covariance $N_o^*/(N_oN_v)\mathrm{Id}_{N_v}$. De plus, la variance minimum d'un estimateur non biaisé peut être reliée à l'équivoque dans l'équation (1.16) : il y a une relation naturelle entre théorie de l'information et mesures statistiques [PFCTPPG06].

Sécurité des techniques substitutives

La méthode de tatouage substitutif de Koch et Zhao [BKZ98] est résistante à WOA : elle est à couverture parfaite. L'attaque KMA permet de trouver la clé ($N_o^* = \log_2 N$), tandis que l'attaque KOA ne permet d'obtenir que les endroits d'insertion, et donc d'écraser le message ($N_o^* = \log_2 N$).

Sécurité des techniques DS

Pour l'approche de Fisher, on obtient les niveaux de sécurité suivants [CFF05] :
KMA : $N_o^* = \sigma_{\mathbf{x}}^2/\psi^2 P$. Le calcul de [CFF05] omet le caractère aléatoire de la clé. Si on le prend en compte, on obtient pour KMA [CPFPG05a] :

$$N_o^* = LJP\frac{\sigma_{\mathbf{x}}^2\sigma_{\mathbf{c}}^2}{\sigma_{\mathbf{c}}^2 + \sigma_{\mathbf{x}}^2/N_o} = \frac{\sigma_{\mathbf{x}}^2}{\psi^2(\frac{1}{PLJ} + P\sigma_{\mathbf{x}}^2/N_o)}$$

WOA : on rajoute l'incertitude liée aux bits du message. Cependant, le niveau de sécurité théorique est le même que pour KMA : $N_o^* = \sigma_{\mathbf{x}}^2/\psi^2 P$. Les porteuses ne sont identifiées qu'à une permutation près entre elles, et au signe près. L'attaquant peut donc altérer le tatouage, mais pas forcément en écrire un autre.
KOA : même si l'on connaît le tatouage, il y a une incertitude liée à la superposition des porteuses. Il faut donc séparer les sources (problème de calcul matriciel). Des algorithmes existent, notamment adaptés au cas gaussien. Là encore, les porteuses ne sont identifiées qu'à une permutation près entre elles, et au signe près. N_o^* est difficile à exprimer mais la trace tr(FIM$(\mathbf{k})^{-1}$) décroît plus vite que $1/N_o$. Le nombre d'observations nécessaires est de l'ordre de $O(LJ)$. L'analyse pratique s'étend à l'ISS et au SCS (pour KOA et WOA). En pratique, la sécurité de ISS dans le cas KOA est aussi basse que celle de DS [CFF05].

Pour l'approche de Shannon, si la clé est générée suivant une loi normale $c_k \sim \mathcal{N}(0, \sigma_{\mathbf{c}}^2)$ [CPFPG05b][CPFPG05a] :

$$\mathrm{I}(\mathbf{y}^1, \ldots, \mathbf{y}^{N_o}; \mathbf{k}|\mathbf{m}^1, \ldots, \mathbf{m}^{N_o}) = \frac{N}{2}\log\left(1 + \frac{N_o\sigma_{\mathbf{c}}^2}{\sigma_{\mathbf{x}}^2}\right)$$

$$H(\mathbf{k}|\mathbf{y}^1, \ldots, \mathbf{y}^{N_o}, \mathbf{m}^1, \ldots, \mathbf{m}^{N_o}) = \frac{N}{2}\log\left(\frac{2\pi e\sigma_{\mathbf{c}}^2\sigma_{\mathbf{x}}^2}{\sigma_{\mathbf{x}}^2 + N_o\sigma_{\mathbf{c}}^2}\right)$$

Si une seule image est tatouée avec N_c différentes clés (tatouage multiple), on obtient selon l'attaque les résultats suivants.

KMA :

$$H(\mathbf{k}^1,\dots\mathbf{k}^{N_c}|\mathbf{y},\mathbf{m}^1,\dots,\mathbf{m}^{N_c}) = \frac{N}{2}\log\left(\left(2\pi e\frac{\sigma_{\mathbf{c}}^2}{N_c}\right)^{N_c}\frac{\sigma_{\mathbf{x}}^2}{\sigma_{\mathbf{x}}^2+\sigma_{\mathbf{c}}^2}\right)$$

Dans [CPFPG05a], $\sigma_{\mathbf{c}}^2 = \frac{1}{\gamma^2 P}$ car la clé n'est pas normalisée.
KOA :

$$I(\mathbf{y};\mathbf{k}^1,\dots,\mathbf{k}^{N_c}) = N\left(\frac{1}{2}\log(2\pi e(\sigma_{\mathbf{x}}^2+\sigma_{\mathbf{c}}^2)) - H(Y|\mathbf{k}^1,\dots,\mathbf{k}^{N_c})\right) \qquad (1.15)$$

Le dernier terme doit être calculé numériquement.
EOA :

$$I(\mathbf{y};\mathbf{k}^1,\dots,\mathbf{k}^{N_c}|\mathbf{x}+\tilde{\mathbf{x}})$$

où $\tilde{\mathbf{x}}$ est l'erreur d'estimation de l'original et doit être calculé numériquement.
Le scénario correspondant à l'attaque d'oracle (cf. paragraphe 1.2.3) conduit à étudier $\mathrm{I}(\mathbf{m}^1,\dots,\mathbf{m}^{N_o};K|\mathbf{y}^1,\dots,\mathbf{y}^{N_o})$ et à l'erreur d'estimation de la clé suivante :

$$\sigma_E^2 \geq \frac{1}{2\pi e}e^{2H(K|\mathbf{y})} \quad . \qquad (1.16)$$

Les conclusions communes à ces deux études sont que le niveau de sécurité de DS dépend de DWR et que toutes les observations fournissent la même quantité d'information : la fuite d'information est linéaire.

Dans [BC06a], on distingue de plus différents types de sécurité pour les algorithmes de type DS. La **sécurité de la clé** consiste à rendre impossible l'estimation des porteuses, même à une permutation près. Dans le cas contraire, l'algorithme est dit **non sûr**. La **sécurité du sous-espace** consiste à rendre impossible l'estimation du sous-espace de tatouage, c'est à dire d'une base orthonormale de l'espace généré par les porteuses. Elle implique la sécurité de la clé. DS et ISS sont donc non sûrs.

Sécurité des techniques quantificatives

La sécurité théorique des techniques quantificatives n'a pas été étudiée par l'approche de Fisher. Avec celle de Shannon, on montre que les fuites d'information sur la clé des méthodes quantificatives sont plus importantes que celles de l'étalement de spectre [PFCPG05b], si la clé est uniquement la séquence d'agitation. Pour DC-DM, dans le cas **EOA :**

$$\mathrm{I}(\mathbf{y}^1,\dots,\mathbf{y}^{N_o};\mathbf{k}|\mathbf{m}^1,\dots,\mathbf{m}^{N_o}) = N\left(-\log(1-\alpha)+\sum_{k=1}^{N_o}\frac{1}{k}\right)$$

avec $H(K) = N\log(\Delta)$. On voit que pour $\alpha = 1$, la fuite d'information est infinie. Intuitivement, cela correspond au fait que la grille de quantification utilisée peut être détectée à chaque observation, donc $(d_k+m_k^{n_o})$. La sécurité de DM doit donc s'appuyer sur une forte compensation des distorsions, par exemple $\alpha = 0.5$, qui, si elle n'est pas justifiée par le présence de $\mathbf{n}$, nuit à la robustesse et à la capacité. Une autre particularité importante de DC-DM est que le niveau de sécurité est pratiquement indépendant de DWR pour des valeurs pratiques [PFCTPPG06].

Si l'on prend en compte l'entropie de la clé, on constate que l'équivoque est bien meilleure pour le schéma de Costa idéal car la taille du dictionnaire croît très rapidement. Des progrès sont donc possibles concernant la sécurité de DC-DM [PFCPG05b].

La sécurité de QIM pourrait également s'appuyer sur le secret de la grille de quantification, et sur la quantification vectorielle à haute dimension. Les pistes proposées consistent à effectuer des rotations de la grille dépendantes d'une clé, à rendre la grille aléatoire, ou à rendre secrète la fonction g d'adaptation du pas de quantification dans une technique RDM. Dans ces deux derniers cas, il est cependant difficile de réellement contrôler la distorsion à l'insertion [PFCTPPG06].

1.4.3 Algorithmes pratiques d'attaques sur la sécurité

Le niveau théorique de sécurité calculé précédemment fournit une borne inférieure de la sécurité pour le tatoueur, pour un algorithme donné, ce qui correspond à une sorte de "débit accessible" du problème de la sécurité. Cependant, cette borne peut se révéler impossible à atteindre calculatoirement et doit être complétée par une étude d'attaques pratiques sur la sécurité [PFCTPPG06] (des tentatives ont été d'ailleurs effectuées pour évaluer la "sécurité calculatoire" d'un algorithme de tatouage [Kat05]).

L'approche de Fisher fournit des pistes pour construire des estimateurs pratiques de la clé. Les algorithmes pratiques d'attaque sur la sécurité de DS correspondants utilisent pour KMA un estimateur du maximum de vraisemblance (ML). Pour KOA, on s'appuie sur des algorithmes de Séparation Aveugle de Source (BSS), car l'énergie du tatouage est concentrée dans un sous-espace. L'Analyse en Composantes Principales (ACP) est utilisée, ou encore l'Analyse en Composantes Indépendantes (ACI), extension de l'ACP qui contraint le résultat à des vecteurs indépendants. Enfin, pour WOA, on se ramène à une BSS dans un environnement bruité. Ces algorithmes sont également efficaces en pratique face à LISS [CFF05] (seule la puissance affectée à la transmission du message sur les porteuses diminue).

Les attaques pratiques sur la sécurité des techniques quantificatives s'appuient une estimation du signal d'agitation fondée sur la théorie des ensembles [PFPGV06]. L'estimation de la transformée d'étalement dans STDM ou SSP est possible grâce à l'algorithme ACI construit précédemment pour l'étalement de spectre [CFF05].

1.4.4 Vers des algorithmes de tatouages plus sûrs

L'étude théorique de la sécurité des techniques de tatouage ayant révélé des failles, il paraît naturel de les corriger. Des algorithmes de tatouage centrés sur le problème de la sécurité ont ainsi été proposés. Cependant, la sécurité apparaît pour l'instant comme un compromis avec la robustesse.

[BC06b] propose une technique adaptée de l'étalement de spectre amélioré appelée "tatouage naturel" (NW, *Natural Watermarking*). Elle consiste à "brouiller" le résultat d'une estimation de la clé par ACI. Sur la *fig.* 1.16, on montre les distributions de $< \mathbf{z}, \mathbf{c}^i >$ en une dimension ($LJ = 1$) et en deux dimensions ($LJ = 2$). Pour DS, les mots de code possibles en 2D sont les points $(-1,+1),(-1,-1),(+1,-1),(+1,+1)$. Or une ACI consiste à trouver, dans les données observées, des directions dont les projections conduisent à des distributions singulières. Dans l'exemple de la *fig.* 1.16, si l'on suppose que le sous-espace a été estimé, la recherche de $\mathbf{c}^1$ et $\mathbf{c}^2$ revient à opérer des rotations sur la base du sous-espace, jusqu'à révéler les quatre singularités. L'estimation est bruitée par les documents hôtes. Pour le tatoueur qui connaît les porteuses, le décodage revient à identifier le quadrant auquel appartient le point. Le tatouage naturel

s'appuie quant à lui sur la règle d'insertion suivante :

$$\mathbf{w} = \sum_i (b_i| < \mathbf{x}, \mathbf{c}^i > |- < \mathbf{x}, \mathbf{c}^i >) \frac{\mathbf{c}^i}{\|\mathbf{c}^i\|^2}$$

Le décodage s'effectue toujours par corrélation. Le tatouage naturel opère donc une insertion informée qui pré-annule les interférences de l'hôte. Elle revient à affecter b_i au signe de $< \mathbf{y}, \mathbf{c}^i >$. Le principe de la technique est de ne pas modifier la distribution de $< \mathbf{x}, \mathbf{c}^i >$: $f_{<\mathbf{x},\mathbf{c}^i>} = f_{<\mathbf{y},\mathbf{c}^i>}$. La distribution conjointe des porteuses est une distribution gaussienne multivariée (cf. *fig.* 1.16). Si les documents hôtes suivent eux-mêmes une distribution gaussienne multivariée, le tatouage naturel assure une sécurité du sous-espace. Dans le cas contraire, le pirate peut identifier un sous-espace où le tatouage suit une distribution gaussienne multivariée. Néanmoins, l'algorithme assure toujours une sécurité de la clé. Le défaut principal du tatouage naturel est son coût en termes de robustesse. En effet, la *fig.* 1.16 montre que les mots de codes les plus probables sont situés près de l'origine. La puissance du tatouage accordée à la transmission de $\mathbf{b}$ sur les porteuses est faible. Un faible bruit additif suffit donc à faire basculer le mot de code dans un autre quadrant. NW est donc moins robuste que ISS, et même DS. NW est l'équivalent pour l'étalement de spectre de la compensation des distorsions dans les techniques quantificatives lorsque $\alpha = 0.5$. Comme dans ce dernier cas, l'augmentation de la sécurité nuit au décodage.

Une variante de cette technique, appelée "tatouage circulaire" (CW, *Circular Watermarking*), est proposée dans [BC06a]. Elle consiste à adopter une distribution du tatouage circulaire, c'est à dire que la distribution conjointe des porteuses ne dépend que d'un scalaire : $f_{\mathbf{c}^1,\ldots,\mathbf{c}^{LJ}} = f(\rho)$. Par analogie avec le signal d'agitation dans les techniques quantificatives, un signal $\mathbf{d}$ est introduit afin de cacher les singularités de la distribution initiale des projections des porteuses. $\mathbf{d}$ suit une distribution circulaire uniforme à L dimensions et est normalisé. La règle d'insertion est la suivante :

$$\mathbf{w}^{n_o} = \sqrt{LJ} \sum_i (\alpha b_i^{n_o} |d_i^{n_o}| \mathbf{c}^i - \lambda < \mathbf{x}, \mathbf{c}^i > \frac{\mathbf{c}^i}{\|\mathbf{c}^i\|}^2)$$

L'insertion revient donc à pondérer dans LISS la part attribuée à la transmission de $\mathbf{b}$ sur les porteuses. Les poids forment un vecteur distribué circulairement, donc ils ne sont pas tous faibles simultanément sur toutes les porteuses et le nombre de "points frontières" diminue (cf. *fig.* 1.16). CW est plus robuste que NW et DS, mais moins robuste que ISS. Sa robustesse diminue avec le nombre de porteuses. CW n'apporte pas de sécurité du sous-espace, car les porteuses sont dépendantes entre elles via $\mathbf{d}$. Seul NW combine une distribution circulaire avec des porteuses indépendantes. CW apporte une sécurité de la clé, car toute rotation de la base dans le sous-espace fournira la même distribution (cf. *fig.* 1.16).

1.5 Contraintes spécifiques au tatouage d'images

Les techniques de tatouage évoquées jusqu'à présent étaient génériques et pouvaient s'appliquer à divers supports : audio, image vidéo... Dans cette thèse, on s'intéressera plus particulièrement au tatouage d'images "naturelles", c'est-à-dire de type photographique et réaliste, largement distribuées sur Internet. Cette définition écarte par exemple les images synthétiques (dessin, image 3D) ou présentant des propriétés

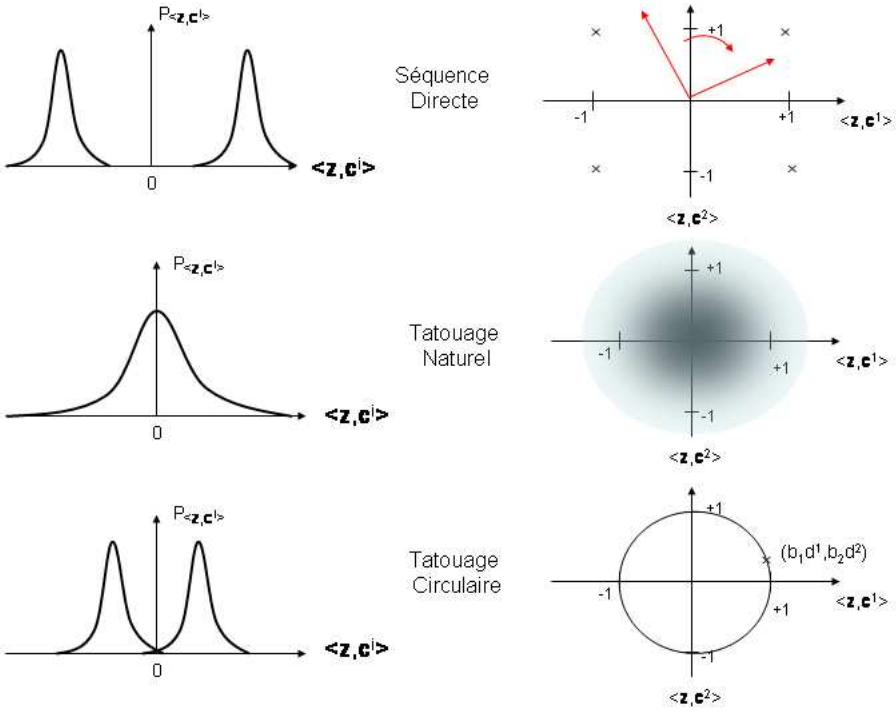

FIG. 1.16 – Distributions des projections sur le sous-espace de tatouage : DS, NW et CW

statistiques particulières dues à leur formation : issues de certains appareils photos, sous certains formats de stockage, images *halftone*, images de télédétection, images médicales... Jusqu'à présent, on a considéré que le document hôte du tatouage était un signal stationnaire, et souvent blanc gaussien. Nous verrons que ce modèle est peu adapté aux images naturelles, qui sont très difficiles à modéliser. L'utilisation d'images détermine également l'utilisation de domaines d'insertion du tatouage appropriés, ainsi qu'une modélisation psychovisuelle destinée à fournir des distances perceptuelles et des masques perceptuels. Enfin, nous présentons les techniques classiques de resynchronisation face aux attaques géométriques d'images, ainsi que les techniques de tatouages exploitant les propriétés statistiques de l'image.

1.5.1 Domaines d'insertion envisageables

On a vu que l'insertion du tatouage peut s'effectuer dans n'importe quel domaine transformé, à condition que l'on respecte la contrainte d'imperceptibilité et que la transformation Tr soit inversible. Dans le cas le plus simple, Tr est la fonction identité. On parle alors de **domaine spatial** : x_{k_1,k_2} est la luminance du pixel (k_1, k_2). Hernandez *et al.* se sont notamment intéressés à ce domaine [HPGRN98]. Les simulations effectuées par la suite concerneront les domaines de la luminance et de la DCT par blocs.

Domaines invariants

Bien qu'elle minimise le coût calculatoire, l'inconvénient d'une insertion dans le domaine spatial est sa sensibilité aux attaques géométriques et à la compression avec perte. C'est pourquoi beaucoup de schémas de tatouage utilisent une insertion dans le

domaine fréquentiel. Par exemple, dans le **domaine de la DFT** (transformée de Fourier discrète en 2D), une translation devient une simple modification de phase. En travaillant sur le module des coefficients, le tatouage est robuste à cette attaque. La difficulté est de trouver un espace complètement invariant à des attaques. O'Ruanaith et Pun [JP98] ont proposé un espace d'insertion invariant aux translations, rotations et changements d'échelle. Ils utilisent la **transformée de Fourier-Mellin**, qui est un changement de variable par une transformation polaire logarithmique suivie d'une transformée de Fourier. Le changement de variable est $(x, y) \rightarrow (\mu, \theta)$, avec $x = e^{\mu \cos\theta}$ et $y = e^{\mu \sin\theta}$). L'implantation de cette méthode est cependant difficile, car il n'existe pas de réelle bijection entre le domaine spatial et le domaine transformé, qui est continu. Il faut donc passer par une interpolation, ce qui nuit aux performances. C'est pourquoi ce domaine, souvent cité à titre d'exemple, est rarement repris en pratique.

Domaine de la DCT

La DCT, ou transformée en Cosinus Discrète, en 2D d'une image $\mathbf{x}$ est définie par :

$$t_{\mathbf{x}}(u,v) \triangleq \frac{2}{\sqrt{N_1 N_2}} K_{u,v} \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} x_{k_1,k_2} \cos\pi \frac{(2k_1+1)u}{2N_1} \cos\pi \frac{(2k_2+1)v}{2N_2}$$

où $K_{u,v} = (\frac{1}{2})^{\delta(u)+\delta(v)}$. Sa transformée inverse est :

$$x_{k_1,k_2} = \frac{2}{\sqrt{N_1 N_2}} \sum_{u=0}^{N_1-1} \sum_{v=0}^{N_2-1} t_{\mathbf{x}}(u,v) K_{u,v} \cos\pi \frac{(2k_1+1)u}{2N_1} \cos\pi \frac{(2k_2+1)v}{2N_2}$$

Cette transformation est souvent calculée sur des blocs de l'image de taille 8x8, soit 64 coefficients. Dans ce cas, on notera les coefficients du domaine transformé $t^k_{\mathbf{x}}(u,v)$, k désignant le bloc et (u,v) correspondant toujours aux fréquences. La propriété essentielle de la DCT est la décroissance rapide de l'amplitude des coefficients lorsque u et v augmentent, ce qui la rend utile pour la compression d'images. De plus, la DCT offre une bonne robustesse à des attaques telles qu'ajustement de brillance et de contraste, correction gamma, filtrage, floutage et compression.

Dans l'article fondateur du tatouage d'image par étalement de spectre, Cox *et al.* [CKLS97] proposent une insertion dans le **domaine de la DCT** appliquée à l'image complète pour une robustesse à la compression avec perte et au changement d'échelle. Leur algorithme a cependant recours à une comparaison avec l'image originale pour extraire le tatouage : il est non aveugle. Le **domaine de la DCT par bloc** est particulièrement populaire dans le domaine du tatouage, par exemple dans [PBBC97] qui est assez similaire à [CKLS97] mais de type aveugle. En effet, l'algorithme de compression JPEG réside en une quantification des coefficients de la DCT par blocs 8x8 selon leur importance perceptuelle. L'utilisation de cette transformation comme domaine d'insertion permet donc une meilleure résistance à la compression JPEG et permet éventuellement d'insérer le tatouage sans décompresser l'image. Le même principe est appliqué au tatouage vidéo avec le format de compression MPEG [Har99], le gain calculatoire étant plus important. Un dernier avantage est de profiter des nombreuses études perceptuelles réalisées pour mettre au point l'algorithme de compression JPEG. L'utilisation de masques et de mesures perceptuels est donc facilitée.

Tatouage dans le domaine des ondelettes (DWT)

De la même manière que pour la DCT, la transformée en ondelettes (domaine temps-échelle) est utilisée en raison de son rôle dans le nouveau standard de compression JPEG-2000 [KH98]. Il est donc plus facile de développer des masques psychovisuels efficaces. L'insertion a lieu dans une sous-bande donnée ou dans des arbres d'ondelettes, qui regroupent les coefficients de différentes sous-bandes correspondant à la même localisation spatiale [WL04]. Outre les propriétés de robustesse communes avec la DCT, le tatouage multirésolution est robuste à un changement d'échelle de facteur 2. De plus, les masques perceptuels sont plus fins et il y a moins d'effets de blocs. On peut donc calculer des décodeurs et détecteurs optimaux (cf. paragraphe 1.5.5) [Bru03]. Une méthode quantificative dans le flux de codage JPEG2000 est étudiée dans [Mee01]. On peut également effectuer des calculs de capacité dans ce domaine [WYA03]. Cependant, beaucoup de paramètres interviennent dans un algorithme de DWT (choix de l'ondelette, taux de compression - nombre de coefficients conservés, taille des blocs…), ce qui rend plus difficile une comparaison objective des performances.

Autres domaines

Outre les DFT, DCT, DWT précédemment évoquées, les transformations inversibles classiques en traitement d'images sont la transformation de Karhunen-Loève (KLT), de Hadamard, de Slant et la décomposition en valeurs singulières (SVD) [Jai89]. Le choix de la transformation utilisée dépend de propriétés telles que la compaction d'énergie. Ces quatre domaines ont été utilisés en tatouage d'images, sans apporter de différence significative de performance et pâtissant d'une étude perceptuelle complexe. Le choix d'un domaine adéquat pour une meilleure robustesse du tatouage à la compression est discuté dans [FKK04]. De manière générale, on cherche une représentation de l'image qui décorrèle le bruit. [MM06b] propose par exemple le domaine NMF, qui semble prometteur à cet égard. De nombreux autres domaines transformés ont été envisagés, sans apporter en pratique une amélioration significative des performances ou des invariances géométriques : transformée de Fourier fractionnelle, transformation de Radon-Wigner, transformée de Laguerre discrète, transformation Mojette (cas particulier de transformée de Radon [Aut02]) et même la phase des images [JDB96].

Enfin, on peut modifier les **propriétés statistiques** de l'image. Dans [GLB06], le kurtosis de coefficients de la transformée en ondelettes est modifié, pour une invariance à des attaques valumétriques. La difficulté est de trouver une technique d'insertion : dans [GLB06], on utilise une technique d'optimisation sous contrainte non-linéaire.

Tatouage couleur

Dans ce qui précède, on a travaillé exclusivement à partir de la luminance d'une image. Celle-ci est souvent préférée à la chrominance car elle est plus robuste (par exemple, à la transformation d'une image couleur en niveaux de gris). De plus, le Système Visuel Humain (HVS) est principalement sensible aux changements d'intensité lumineuse. Comme le HVS est moins sensible aux changements de couleurs, les attaques délibérées ou non (compression) sont plus faciles dans la chrominance. On peut cependant procéder au tatouage d'image couleur de plusieurs manières :

- en tatouant dans le domaine de la luminance puis en revenant dans la chrominance

- en concaténant les trois canaux (on a alors $3N$ échantillons au lieu de N, ce qui augmente la capacité et la robustesse)
- en pondérant les canaux pour privilégier le bleu qui est le moins sensible (soit avec $3N$ échantillons, soit avec N échantillons répétés sur chaque canal)

Dans les simulations proposées dans ce document, on utilisera la première technique en cas d'expérimentation sur une image couleur.

Inversement, la faible sensibilité du HVS aux changements de couleurs permet de générer des masques perceptuels plus efficaces. Des filtres améliorent ainsi l'imperceptibilité en tenant compte de la sensibilité de l'œil à chaque canal. Des artefacts peuvent cependant apparaître et les méthodes privilégiant une couleur précise sont plus délicates à mettre en œuvre. On peut aussi utiliser un espace achromatique afin de prendre en compte les spécificités perceptuelles des trois composantes couleurs [DP04].

1.5.2 Modèles d'images naturelles

La littérature de référence en traitement d'images fournit peu d'informations sur des modèles simples d'images naturelles [Mai02][Jai89][CP95].

Modèles statistiques

En l'absence de modèle plus approprié ou par souci de généralité, on a souvent recours au **modèle gaussien**, notamment dans le domaine spatial. Certains auteurs modélisent l'image par des canaux gaussiens parallèles. Dans le domaine transformé (DCT par blocs ou DWT), le **modèle gaussien généralisé** (GGD) convient [HAPG00]. Sa densité de probabilité f_x est définie par

$$f_x(x) = Ae^{-|\beta x|^c} \ ,$$

A et β étant calculés à partir de la fonction gamma notée Γ, de la variance σ et d'un paramètre de courbure c :

$$\beta = \frac{1}{\sigma}\left(\frac{\Gamma(3/c)}{\Gamma(1/c)}\right)^{1/2} \qquad A = \frac{\beta c}{2\Gamma(1/c)} \ .$$

$c = 1$ correspond à une loi de Laplace et $c = 2$ à une gaussienne. Ces distributions sont également connues sous le nom de distributions de Subbotin (cf. [JKB95], pp. 195-197, avec $c = 2/\delta$, $\alpha = (\Phi 2^{1/c})^{-1}$ et en utilisant $\Gamma(x+1) = x\Gamma(x)$). Les GGD sont utilisées pour modéliser une grande variété de bruits et de phénomènes aléatoires dans divers contextes : acoustique, technologie video. Leur utilisation en traitement d'images provient du codage d'images, où les coefficients de la DCT ont d'abord été considérés comme laplaciens avant qu'on étudie différentes valeurs du paramètre de courbure c [JF95][F. 93][RG83]. Il est également possible de modéliser les coefficients des sous-bandes de la DWT par différentes distributions gaussiennes généralisées [Mal89][DM03]. Dans cette thèse, nous utiliserons le modèle GGD dans les parties 1.5.5 et 4.4.

Chacun des 64 coefficients $t_k(u,v)$ de la DCT par blocs 8x8 suit une GGD :

$$f_x(x) = A_{u,v}e^{-|\beta_{u,v}x|^{c_{u,v}}} \ ,$$

$A_{u,v}$ et $\beta_{u,v}$ étant calculés à partir de la fonction gamma, de la variance $\sigma_{u,v}$ et d'un paramètre de courbure $c_{u,v}$. $c_{u,v}$ et $\sigma_{u,v}$ peuvent être estimés pour chaque coefficient

et chaque image car chaque bloc $t^k_{\mathbf{x}}(u,v)$ constitue une réalisation. Le plus souvent, le paramètre $c_{u,v} = 0.8$ offre de bons résultats pour les moyennes fréquences. Cette modélisation a été utilisée notamment dans [HPG99] afin de calculer une stratégie de détection optimale. Elle est courante dans les domaine du tatouage par quantification où les performances sont le plus souvent calculées à l'aide de simulations de Monte-Carlo [EBTG03][PGBM03][BBP04].

Dans le même esprit, Barni *et al.* [BBPR98] modélisent chacun des N^2 coefficients de la DCT calculée sur l'image entière par une loi de Laplace. Cette fois-ci, les paramètres sont estimés à partir d'un ensemble de 170 images test. Enfin, Barni *et al.* [BBRP00] modélisent le module des coefficients de la DFT sur toute l'image par une loi de Weibull :

$$f_x(x) = \frac{\beta}{\alpha}\left(\frac{x}{\alpha}\right)^{\beta-1} e^{-(x/\alpha)^\beta}$$

$\beta = 1$ correspond à une loi exponentielle et $\beta = 2$ à une loi de Rayleigh. Là encore, chaque image correspond à une réalisation.

[SLSZ03] fournit des modèles pour les densités marginales d'une image (après filtrage), ainsi que des modèles inspirés de la physique (superposition, occlusion). On définit la dérivée horizontale d'une image comme la différence entre deux pixels adjacents. Cette dérivée suit une loi gaussienne généralisée [LMH01] de courbure $c = 0.68$ sur une base d'image test. La combinaison linéaire de plusieurs pixels adjacents tend à suivre une distribution laplacienne généralisée [MSS05][Gre02]. Cette propriété avait déjà été observée dans [HM99]. Cette propriété sera utile dans la partie 4.4.

Modèles de corrélation spatiale

Un modèle courant pour une ligne d'une image naturelle est le processus stationnaire de Markov du premier ordre [Jai89]. Sa fonction de covariance est $K_{\mathbf{x}}(n) = \rho^{|n|}, \rho < 1, \forall n$. Sa matrice de covariance $K_{\mathbf{x}}(m,n) = K_{\mathbf{x}}(m-n)$ a une structure de Toeplitz, comme pour tout processus stationnaire. Ce modèle n'est cependant pas suffisant. Le modèle des champs de Markov est souvent utilisé pour décrire les textures d'une image [Mai02][CP95].

Modèles utilisés en stéganalyse

La stéganalyse, pendant de la stéganographie, a pour but de détecter un tatouage par la modification qu'il introduit dans les propriétés statistiques d'une image. Elle apporte donc des modèles d'images intéressants. Les moments d'ordre 1 à 4 du signal sont souvent utilisés [MSS05]. Martin *et al.* évoquent entre autres une modélisation de la dérivée horizontale d'une image par la superposition d'un nombre aléatoire de répétitions du même objet [MSS05], proche du modèle d'occlusion décrit dans [SLSZ03]. Sullivan *et al.* [SMCM05] modélisent le signal par une chaîne de Markov. L'effet du tatouage est visible sur la matrice de co-occurrence (estimation de la densité de probabilité). Le tatouage affaiblit les corrélations au sein de l'image, ce qui provoque un étalement autour de la diagonale. Nous reprendrons ce modèle dans l'annexe C.4.

1.5.3 Imperceptibilité : distances et masques perceptuels

La qualité perceptuelle devrait être évaluée par des expériences subjectives avec des observateurs humains. Celles-ci sont cependant coûteuses et rares. La contrainte d'imperceptibilité conduit donc à construire des critères objectifs pour mesurer l'impact per-

ceptuel d'un tatouage : ce sont les "mesures perceptuelles". Dans le cadre du tatouage d'images, elles sont construites à partir d'études psychovisuelles et de modèles du Système Visuel Humain (HVS). On parle également de "fidélité" pour signifier qu'il y a peu de différences entre une image modifiée et l'image originale (par exemple en erreur quadratique moyenne (EQM)), par opposition à "qualité", qui désigne l'attrait perceptuel de l'image modifiée [CMB02]. On préférera ici parler de "qualité perceptuelle". Les mêmes études psychovisuelles permettent l'élaboration de tatouages maximisant la qualité perceptuelle, grâce à l'usage de "masques psychovisuels".

Distances perceptuelles

Critères statistiques : le DWR et le WNR, utilisés pour contrôler la puissance du tatouage et de l'attaque, sont des critères de qualité perceptuelle, qui sont liés à l'EQM. Cependant, ces critères conduisent à surestimer certaines distorsions, telles que les distorsions géométriques. Notons que le WNR reste un bon estimateur de la puissance de l'attaque, donc de son impact sur un algorithme de tatouage. On a déjà défini le DWR :

$$\text{DWR} \triangleq \frac{\sum_{k=1}^{N} x_k^2}{\sum_{n=1}^{N} w_k^2}$$

L'Erreur Quadratique Moyenne (EQM) ou *Mean Squared Error* (MSE) est définie par :

$$\text{EQM}_I \triangleq \text{E}[(\mathbf{y}-\mathbf{x})^2], \quad \text{EQM}_A \triangleq \text{E}[(\mathbf{z}-\mathbf{y})^2]$$

EQM_I désignant l'EQM d'insertion et EQM_A l'EQM d'attaque. En traitement d'images, on utilise la notion de PSNR (*Peak Signal to Noise Ratio*) qui prend en compte la plus grande valeur possible du signal notée max et l'EQM [DP04] :

$$\text{PSNR}_I \triangleq \frac{max^2}{\text{EQM}_I}, \quad \text{PSNR}_A \triangleq \frac{max^2}{\text{EQM}_A} \ .$$

Pour une image codée sur 8 bits par pixel et une insertion dans la luminance, max=255. Si le modèle de tatouage ou de bruit est additif, pour une image $\mathbf{x}$ donnée, l'EQM, le PSNR et le DWR ou WNR sont liés par un facteur multiplicatif (additif en dB) dépendant de $\mathbf{x}$. Dans la suite de cette thèse, on utilisera donc uniquement le DWR.

On considère généralement en tatouage d'images qu'un tatouage est imperceptible pour PSNR> 36 dB. Sur les images utilisées dans les simulations de la suite de cette thèse, nous utiliserons généralement des valeurs plus restrictives. Par exemple, DWR=28 dB correspondra à PSNR=43,5 dB en moyenne.

La distance de Kullback-Leibler (ou entropie relative) peut évaluer la distance statistique entre le document hôte et le document tatoué. La distance de Kullback-Leibler entre variables aléatoires X et Y de distributions respectives f_X et f_Y est [Bla87] :

$$D_{\text{KL}} \triangleq \int f_X \log_2 \left(\frac{f_X}{f_Y} \right)$$

Les rôles de X et Y n'étant pas symétriques, il ne s'agit pas d'une vraie métrique. Cependant, il s'agit d'une fonction convexe, non négative, nulle si et seulement si $X = Y$.

Critères psychovisuels : on distingue les métriques à référence complète, qui font appel à la carte d'erreur entre l'image originale et l'image altérée ; à référence réduite,

lorsque seules des descriptions des images, comme des points d'intérêts, sont nécessaires ; et sans référence, lorsque seule l'image altérée est connue.

La distance perceptuelle fondée sur des critères psychovisuels la plus courante est la distance de Watson [Wat93]. La distance repose sur le calcul d'un "seuil de visibilité" et d'un "effet de masque de contraste". Le seuil de visibilité, est également appelé Différence Juste Perceptible (JND pour *Just Noticeable Difference*). Un dépassement du JND est censé être repérable par 50% des observateurs humains [CMB02]. Selon le modèle de Watson *et al.*, le seuil de visibilité est l'amplitude maximale $t_m^k(u,v)$ d'une altération invisible d'un coefficient donné $t_{\mathbf{x}}^k(u,v)$ de la DCT. Il est calculé à partir des fréquences spatiales $f_{u,v}$ (en cycle par degré) associées à chaque coefficient, qui dépendent de la taille d'un pixel sur l'écran et de la distance de vision. Leur influence perceptuelle a été déterminée par des expériences psychovisuelles effectuées à partir de mires sinusoïdales d'une fréquence et d'une orientation donnée, qui ont par la suite été utilisées dans la norme JPEG [AP92][SWA94]. Le seuil est pondéré pour chaque bloc par l'intensité du coefficient DC (effet d'intensité d'arrière-plan). Le masque de contraste détermine le contraste maximum au sein d'un motif lorsqu'il est superposé de façon invisible à un autre motif masquant d'un contraste, d'une fréquence spatiale et d'une orientation données [DDML02]. Finalement, on mesure l'erreur perceptuelle entre une image de référence $\mathbf{x^1}$ et une image modifiée $\mathbf{x^2}$:

$$\epsilon_k(u,v) \triangleq \frac{|t^k_{\mathbf{x^1}}(u,v) - t^k_{\mathbf{x^2}}(u,v)|}{t_m^k(u,v)}$$

est l'erreur dans la DCT exprimée en JND puis,

$$\epsilon(u,v) \triangleq \left(\sum_k |\epsilon_k(u,v)|^{\beta_s}\right)^{1/\beta_s}, \quad D_W \triangleq \left(\sum_{u,v} \epsilon(u,v)^{\beta_f}\right)^{1/\beta_f}$$

sont respectivement l'erreur dans la bande fréquentielle (u,v) et l'erreur totale, avec en général $\beta_s = \beta_f = 4$ dans la distance de Minkowski. Watson *et al.* ont effectué le même type d'étude pour les matrices de quantification de la transformée en ondelettes [WYSV97].

Avcibas *et al.* analysent dans [ASS02] différentes mesures de qualité d'image et appliquent leurs travaux à la stéganalyse. La thèse de Jean-Luc Olivès [Oli98] offre une comparaison de nombreux critères psychovisuels. En particulier, il souligne les bonnes performances de la méthode *Visible Differences Predictor* de S. Daly [Dal94], utilisant le domaine spatial, qui offre une carte 2D des déformations visibles sur une image. Lambrecht et Farrell [vdBLF96] proposent une autre mesure perceptuelle pour les images couleur, utilisant les bancs de filtres de Gabor (réglés selon différentes fréquences spatiales et orientations, ce qui donne une sorte de représentation multirésolutions). Son avantage est de ne pas diviser l'image en blocs. Dans le domaine du tatouage couleur, on trouve également des mesures perceptuelles spécifiques dans l'espace chromatique uniforme S-CIELAB, dont le principe est de séparer les influences de l'illumination et du contraste.

Une nouvelle approche est fondée sur les distorsions structurelles [WBSS04]. La mesure perceptuelle, appelée SSIM (*Structural SIMilarity*), effectue une corrélation entre les motifs des images, indépendamment de la luminance locale (qui est soustraite) et du contraste (qui est normalisé). SSIM est présenté comme une amélioration de deux métriques appelées UQI et RRIQA. C4 est une métrique à référence réduite fondée elle aussi sur l'information structurelle [CCB03]. La métrique du Gain Composé (*Compound Gain*) est une généralisation de la distance de Kullback-Leibler, calculée autour

des points d'intérêt de l'image [GFVFVRS01]. Enfin, Komparator est une métrique proche du *Visible Differences Predictor* et étendue aux images couleur [BC03].

Dans [MACC07], différentes métriques sont comparées à une évaluation subjective rigoureuse dans l'application spécifique au tatouage. Les auteurs concluent que les mesures C4 et surtout Komparator sont fiables. *A contrario*, les métriques UQI, RRIQA, SSIM et le PSNR ne sont pas fiables dans le cas particulier de l'application au tatouage. Pour nuancer cette conclusion, notons cependant que les algorithmes de tatouage testés dans [MACC07] ne font pas partie des algorithmes classiques (DS, ISS, SCS...) utilisés dans cette thèse.

Limitations des mesures psychovisuelles : la majeure partie des critères de qualité perceptuelle présentés ici est fondée sur des hypothèses communes sur les canaux visuels du HVS et négligent notamment les interactions entre ces canaux, ou entre leurs coefficients. C'est par exemple le cas lorsqu'on somme l'erreur perceptuelle de chaque coefficient dans la distance de Watson. L'efficacité de telles mesures par rapport à des critères plus simples mais notoirement insatisfaisants (MSE, PSNR) est mise en doute [WBL02]. En particulier, il a été montré que les performance de la plupart des modèles sont statistiquement équivalentes à celles du PSNR et de l'EQM [WBSS04]. Aucune mesure de qualité perceptuelle n'est donc totalement satisfaisante.

Les Figs. 1.17 et 1.18 illustrent l'intérêt des mesures de qualité perceptuelle. La *fig.* 1.17 représente la puissance $\sigma^2_{\mathbf{n}}$ (et donc l'impact au décodage) de différentes attaques : bruit AWGN avec DNR entre 12 et 50 dB, compression JPEG de facteur de qualité entre 15% et 100%, rotation puis resynchronisation entre 0° et 5°, translation puis resynchronisation entre (1,1) et (1.5,1.5), filtrage de Wiener d'un bruit entre 15 et 50 dB. On observe entre autres que le DNR introduit par la rotation est très faible et indépendant de l'angle. La *fig.* 1.18 compare les distances SSIM et D_W selon la puissance de chaque attaque. Selon le critère SSIM, un bruit additif gaussien fort déforme fortement l'image, alors qu'un bruit faible est peu visible. Les attaques géométriques resynchronisées (translation, rotation) introduisent un bruit d'interpolation qui a peu d'impact visuel. La faible compression JPEG et le filtrage de Wiener déforment plus l'image que le bruit AWGN. Selon la distance de Watson, ce sont à l'inverse les attaques géométriques qui dégradent le plus l'image. Logiquement, la faible compression JPEG, qui utilise les expériences de Watson *et al.*, est considérée comme très peu perceptible. Le filtrage de Wiener a la même influence qu'un bruit AWGN. Les deux mesures ont donc des résultats très différents. Pour un observateur humain, les déformations les moins visibles seraient dues aux attaques géométriques et à la compression JPEG, tandis que le filtrage de Wiener et l'AWGN sont très gênants. Aucun critère n'étant totalement satisfaisant, l'évaluation perceptuelle dans ce rapport de thèse combinera les résultats des distances DWR, Kullback-Leibler, Watson et SSIM.

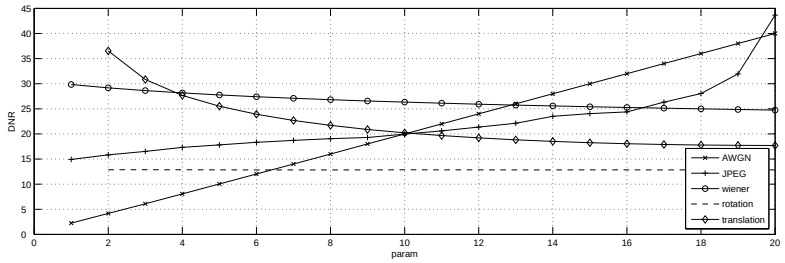

FIG. 1.17 – Comparaison de la puissance de différentes attaques

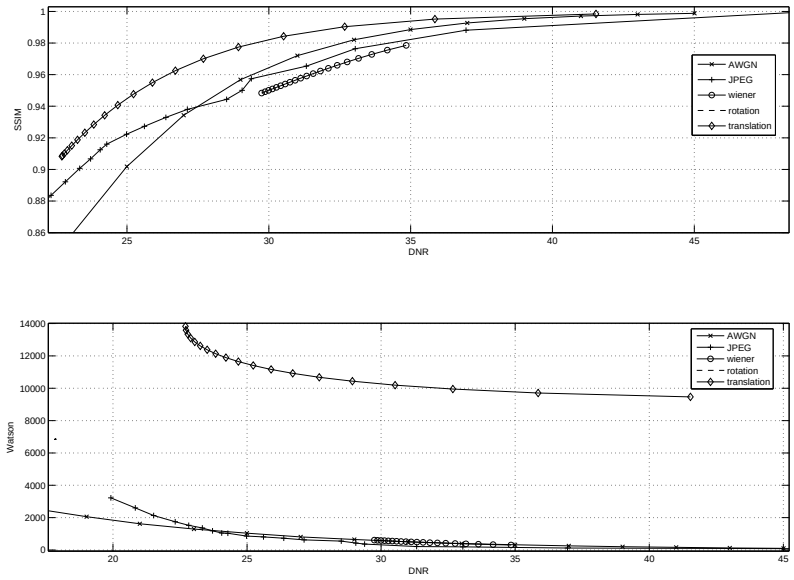

FIG. 1.18 – Comparaison des distorsions perceptuelles de chaque attaque en fonction de sa puissance : (haut) SSIM (bas) distance de Watson

Masques psychovisuels

La contrainte d'imperceptibilité oblige le tatoueur à insérer un tatouage suffisamment faible et dans les composantes les moins perceptibles. Ce compromis pousse à utiliser un masque perceptuel Ψ, calculé en fonction de $\mathbf{x}$, qui pondère $\mathbf{w}$ ou le substitue par un seuil dans l'approche JND ($w_k \leftarrow \psi_n w_k/|w_k|$). Les masques sont souvent inspirés des mesures perceptuelles. La JND impose parfois une contrainte trop stricte sur l'énergie du tatouage, non compatible avec la robustesse. Ψ joue également un rôle dans la robustesse et la sécurité. Par exemple, le spectre 2D d'une image naturelle est anisotrope : il présente des valeurs plus fortes dans certaines directions, selon les contours et textures de l'image. Un tatouage respectant ces propriétés spectrales sera plus robuste au débruitage ou à la compression [VHBP99]. Ainsi, le masque NVF avant modulation par $\mathbf{c}$ et $\mathbf{m}$ possède les mêmes propriétés spectrales que l'image d'origine.

Si l'on pondère par $\Psi = |h_\Psi * \mathbf{x}|$, le masque n'introduit pas de changement de signe. Si P est suffisamment grand, le masque n'interfère donc pas sur le décodage par corrélation ([CMB02], p.227).

Si $\Psi = |h_\Psi * \mathbf{x}|$, on introduit des changements de signe. On doit donc estimer à nouveau le masque avant décodage pour annuler son influence. Pour DS, on calcule alors une corrélation de $\mathbf{z}$ par $\mathbf{c}/\Psi'$, où $\Psi' = |h_\Psi * \mathbf{z}|$. $\mathbf{x}$ et $\mathbf{z}$ étant supposées perceptuellement proches, Ψ' est supposé très proche de Ψ.

Enfin, la plupart des masques spatiaux effectuent un filtrage passe-haut. Celui-ci a souvent pour effet secondaire de stationnariser l'image, à la manière d'un préfiltrage de Wiener avant décodage (cf. paragraphe 1.5.5). Notamment, le masque laplacien (cf. paragraphe 1.5.3) correspond à une annulation des moyennes locales. Dans ce cas, on peut annuler l'influence du masque sur le signe de $\mathbf{w}$ en effectuant une corrélation par $\Psi'\mathbf{c}$. Celle-ci améliore la robustesse au bruit de l'hôte.

Dans cette thèse, on utilisera les masques NVF, Laplacien, d'Alvarez *et al.* et d'Ahumada *et al.* définis ci-dessous.

Masques spatiaux : dans le domaine spatial, l'étalement de spectre utilise toutes

les composantes fréquentielles de l'image, ce qui peut être considéré comme un masque psychovisuel très basique. On peut donc se contenter d'une pondération par un facteur de masquage ψ qui limite la puissance de **w**. Il est cependant préconisé d'utiliser un masque mesurant les variations locales de luminance, car l'œil est moins sensible aux modifications d'amplitude situées près des contours de l'image. Ce modèle psychovisuel très simple s'appelle "loi de Weber" : la sensibilité du HVS est inversement proportionnelle à l'intensité lumineuse [LS04]. Pour un même impact perceptuel, une règle empirique (pas toujours vérifiée) énonce que la puissance maximale de **w** acceptable sous la contrainte d'imperceptibilité est double avec ces masques, soit un gain de 3 dB sur le DWR. L'un des masques spatiaux les plus courants est appelé "filtre Laplacien" (cf. *fig.* 1.19), car il annule les dérivées secondes horizontales, verticales et diagonales de l'image [KJ99] :

$$\psi_{k_1,k_2} = \mathbf{h}_\psi * \mathbf{x}(k_1,k_2)$$

$$h_\psi(k,l) = \frac{1}{9}\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$$

Alvarez-Rodriguez et Pérez-Gonzalez préconisent d'utiliser le filtre [ARPG02] :

$$\psi_{k_1,k_2} = \psi_0 e^{-0.014 M_{k_1,k_2}}$$

$$\text{où } M_{k_1,k_2} = h_\psi(k_1,k_2) * (|m^H_{k_1,k_2}| + |m^V_{k_1,k_2}|)$$

$$\text{avec } \begin{cases} m^H_{k_1,k_2} = x_{n_1+1,n_2} - x_{n_1-1,n_2} \\ m^V_{k_1,k_2} = x_{k_1,k_2+1} - x_{k_1,k_2-1} \end{cases}$$

$$\text{et} \qquad h_\psi(k_1,k_2) = \begin{bmatrix} 0.35^{\sqrt{2}} & 0.35 & 0.35^{\sqrt{2}} \\ 0.35 & 1 & 0.35 \\ 0.35^{\sqrt{2}} & 0.35 & 0.35^{\sqrt{2}} \end{bmatrix}$$

La Fonction de Visibilité du Bruit (NVF) est calculée à partir de la variance locale [VHBP99]. Ce masque insère donc le tatouage dans les régions de l'image correspondant aux textures ou aux contours. La NVF présentée ici utilise un modèle gaussien non stationnaire. Si $\sigma^2_{\mathbf{x}}[k_1,k_2]$ est la variance locale de l'image (par exemple calculée sur des fenêtres 3x3) et $\theta = \frac{D}{\max(\sigma^2_{\mathbf{x}}[k_1,k_2])}$, avec $D \in [50,100]$ réglé expérimentalement, on a :

$$\psi_{k_1,k_2} = \sigma^2_{\mathbf{w}}\left(1 - \frac{1}{1+\sigma^2_{\mathbf{x}}[k_1,k_2]\theta}\right)$$

$\psi \simeq 0$ dans les régions planes, donc on peut augmenter la puissance du tatouage dans ces zones, quitte à dépasser le seuil de visibilité en introduisant les poids S_1 et S_2 :

$$\psi_{k_1,k_2} = S_1\left(1 - \frac{1}{1+\sigma^2_{\mathbf{x}}[k_1,k_2]\theta}\right) + S_2\frac{1}{1+\sigma^2_{\mathbf{x}}[k_1,k_2]\theta}$$

Masques fréquentiels : les masques spatiaux concentrent le tatouage sur les textures et les contours de l'image. Cependant, les contours d'une image concernent peu de points, ce qui réduit la capacité du support et la robustesse. De plus, une modification d'un contour peut générer des artefacts perceptibles. Cette limitation est mise en avant par Delaigle [DDML02] qui préconise l'utilisation de filtres de contraste et de motif, beaucoup plus complexes et faisant intervenir le domaine fréquentiel. Dans le domaine fréquentiel, les masques sont beaucoup plus efficaces et indispensables afin de

ne pas modifier les basses fréquences (composantes les plus perceptibles) ou les hautes fréquences (composantes les plus vulnérables aux attaques). Cox *et al.* proposaient simplement d'utiliser les 1000 coefficients de plus grande amplitude de la DCT [CKLS97]. L'élaboration de masques perceptuels de type JND dans les domaines spatiaux, DCT et ondelettes, inspirés d'autres modélisations du HVS, a fait l'objet des travaux de thèse de F. Autrusseau [Aut02]. Dans [Koz02], on exploite le phénomène de "fovéation" de l'œil humain pour insérer un tatouage dans la périphérie (l'œil s'intéressant plus au centre de l'image) des points d'intérêt de l'image.

Un masque fréquentiel courant utilise le modèle psychovisuel proposé par Ahumada *et al.* Il est notamment appliqué dans une version simplifiée par Hernandez *et al.* [HAPG00]. Miller *et al.* précisent bien que cette mesure n'est pas parfaite, le masque correspondant introduisant par exemple des effets de bloc (on ne travaille que sur un bloc donné de la DCT). De plus, les blocs comportant des contours abrupts contiennent beaucoup d'énergie sur toutes les fréquences, ce qui est confondu avec des blocs très texturés pouvant mieux masquer le tatouage. La *fig.* 1.22 montre l'application de ce masque à l'image Lena. On utilisera ce masque dans les algorithmes DS-DCT (1.2.5) et PCC-DCT (2.1.2), couplé à la sélection de 22 coefficients des moyennes fréquences sur les 64 coefficients de chaque bloc : $\{t^k_{\mathbf{x}}(u,v)| \quad 5 \leq (u+v) \leq 8\}$.

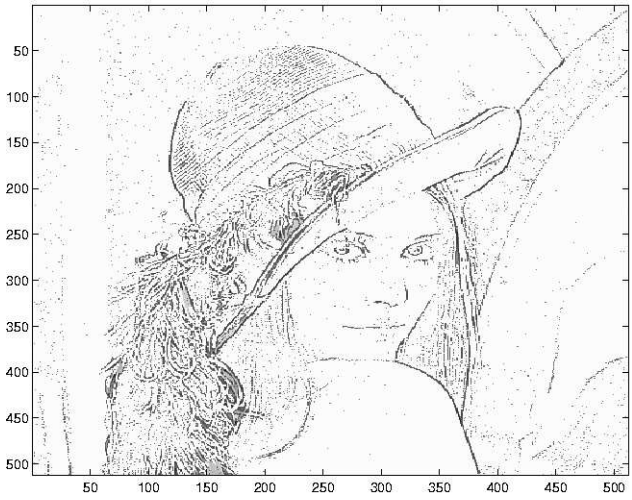

FIG. 1.19 – Exemple de masque de contour (laplacien)

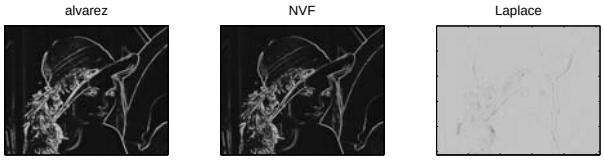

FIG. 1.20 – Comparaison des trois masques de contour (Lena)

1.5.4 Techniques de resynchronisation

Les méthodes de resynchronisation présentées dans cette partie concernent principalement les transformations géométriques affines globales. Les transformations géométriques non affines ou locales peuvent être approchées par une juxtaposition de trans-

FIG. 1.21 – Comparaison des trois masques de contour (Babouin)

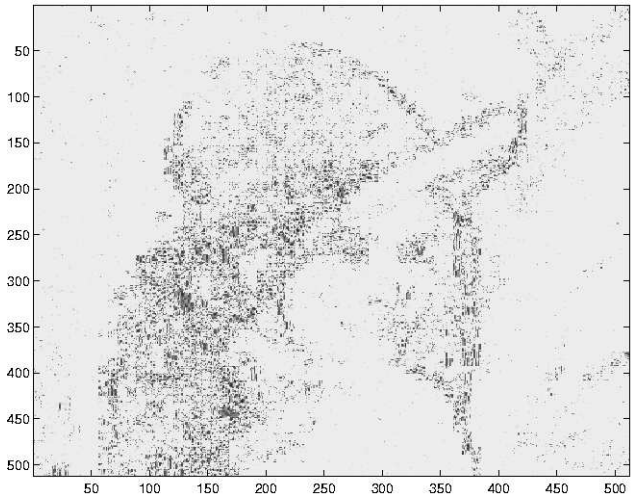

FIG. 1.22 – Exemple de masque psychovisuel calculé dans le domaine de la DCT

formations affines locales [VDP01]. Cependant, ce type d'attaque est très difficile à resynchroniser. Ce problème reste très ouvert pour la communauté du tatouage.

Transformation invariante

La solution la plus élégante au problème de la désynchronisation consiste à insérer **w** dans un domaine invariant aux distorsions affines. Ainsi, l'attaque sera tout simplement ignorée par le décodeur. Dans le domaine de l'image, la transformée de Fourier-Mellin (FMT) [JP98] offre les propriétés les plus intéressantes : invariance aux translations, rotations et changements d'échelle. Cependant, elle reste peu utilisée à cause des difficultés d'implantation : la FMT impose le passage à une représentation continue de l'image et il n'y a donc pas de bijection entre le domaine spatial et l'espace transformé. On doit donc passer par un rééchantillonnage avec perte. Une variante plus facile à implanter consiste à insérer le tatouage sur la projection 1D de l'image transformée par la FMT sur l'axe log-radial [LWB$^+$01]. Plus récemment, un algorithme basé sur la quantification des moments de Zernike d'une image a été proposé [XLP04]. La recherche d'une transformation invariante à l'ensemble des attaques géométriques possibles semble cependant utopique.

Tatouage adapté au contenu

Ce type de méthode utilise des caractéristiques du document propres à son contenu lui-même. Celles-ci sont encore extractibles après attaque, puisque celle-ci ne doit pas affecter la sémantique du document. Notamment, Bas *et al.* proposent d'utiliser les points saillants d'une image [BCD99]. Les inconvénients de ces méthodes sont la nécessité d'une extraction robuste des caractéristiques, et un rapport lointain avec les méthodes classiques (DS ou QIM) les plus étudiées de façon théorique.

Les caractéristiques du document peuvent également servir à la resynchronisation dans des variantes de l'algorithme DS. Par exemple, on peut découper une image en triangles prédéfinis (triangulation de Delaunay) dont les sommets sont des points saillants [BCM00]. Le tatouage est constitué de triangles rectangles de base, remplis par un motif pseudo-aléatoire. Ces triangles sont ramenés à la forme des triangles issus de la triangulation par transformation linéaire et interpolation. Le schéma de tatouage est additif. Au décodage, on calcule à nouveau les points saillants, on ramène les triangles à leur forme de base et on effectue une corrélation. L'insertion peut également avoir lieu dans le domaine fréquentiel : il suffit de calculer la DCT de chaque triangle de base symétrisé. La triangulation ou le passage du triangle quelconque au triangle rectangle peuvent faire appel à une interpolation, mais il s'agit toujours ici d'une détection par corrélation avec une séquence pseudo-aléatoire. La triangulation peut être remplacée par une structure en mosaïque (*tesselation*). Plus généralement, ce type de méthode (extraction des points caractéristiques ou segmentation, formation des motifs élémentaires...) est analysé dans [CSST01].

Insertion d'un motif de resynchronisation

Cette technique consiste à insérer en plus de $\mathbf{w}$ un motif appelé *template* ou "mire" qui permet d'identifier la transformation affine effectuée. Dans [PP03], il s'agit de pics insérés dans les moyennes fréquences de la TFD 2D d'une image. La position de ces pics après transformation affine permet de retrouver la transformation et de l'inverser. Cette technique a été considérablement améliorée dans [LL05], où le motif de resynchronisation inséré tient compte de la connaissance de $\mathbf{x}$ à l'insertion, selon le principe du tatouage informé. Chaque bloc du document est projeté sur la séquence pseudo-aléatoire, et la mire vise à atteindre une région donnée de détection des pics, plutôt que d'être insérée additivement de manière aveugle. Les pics d'autocorrélation sont ainsi plus robustes. Cependant, il reste possible que le pirate s'attaque à la mire elle-même en éliminant les pics de manière aveugle. Les transformations locales ne sont pas non plus prises en compte. Si la mire est commune à plusieurs images tatouées, la méthode est vulnérable aux attaques de collusion (particulièrement pour la vidéo). Une amélioration de la sécurité des motifs de resynchronisation a été proposée dans [DM06], où le motif dépend des propriétés locales de l'image.

Insertion d'un motif périodique

Cette méthode est un cas particulier de mire où le signal synchronisant est confondu avec le tatouage. Elle a été introduite par M. Kutter puis reprise dans une étude théorique par Alvarez-Rodriguez et Pérez-Gonzalez [ARPG02]. Elle consiste à insérer un tatouage appelé *pilote*, présentant des propriété d'autocorrélation, qui consiste en plusieurs (par exemple 4) versions "intercalées" du même motif de base. Intercalé signifie ici imbriqué, avec un décalage horizontal et vertical en image. Des pics (9 ici) sont donc présents dans l'autocorrélation du motif (cf. *fig.* 1.23). Ils subissent les mêmes transformations que le document tatoué. On peut détecter ces pics sans posséder la clé, et sans appliquer de désétalement. Les auteurs proposent d'estimer les paramètres de la transformation en minimisant, par une méthode des moindres carrés non linéaire, la distance entre les pics estimés et une transformation des pics initiaux. Pour la rotation et le changement d'échelle, il subsiste une erreur de quantification sur l'estimation du paramètre, qui est uniformément répartie. En effet, une modification d'un pic n'est détectable que s'il passe d'un pixel à un autre. On ne peut détecter qu'une modification

suffisamment importante, et par conséquent les attaques les plus efficaces sont les plus légères : elles introduisent au moins un bruit d'interpolation. De plus, la technique n'est pas robuste à l'élimination de la mire (par exemple par filtrage à moyenne mobile pour modifier les pics) et l'attaque Stirmark (combinaison d'attaques locales).

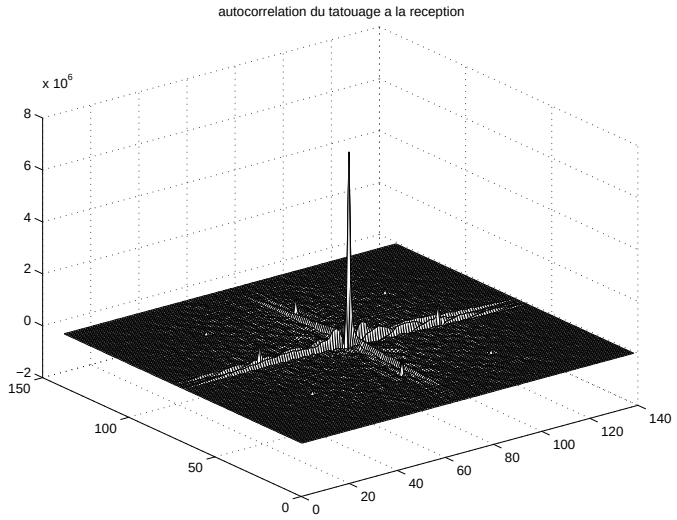

FIG. 1.23 – Autocorrélation au décodage, $\theta = 5^{\circ}$, $\delta_h = \delta_v = 43$, DWR=18 dB

Recherche exhaustive

La recherche exhaustive consiste à appliquer la détection ou le décodage à l'ensemble des versions transformées de la clé. Par exemple, on effectue la corrélation avec une transformation affine du vecteur code dont les paramètres varient dans un intervalle donné [HPGRN98]. Cette méthode est souvent écartée en raison de son coût calculatoire qui augmente exponentiellement avec la taille de l'espace de recherche. Cependant, M.Barni montre qu'en théorie la détection par recherche exhaustive (*Exhaustive Search Detection*, ESD), tout comme la détection par mire (*Template Matching Detection*, TMD), sont efficaces si la taille de l'espace de recherche augmente de façon polynomiale, à condition d'augmenter la redondance ou la taille de la mire en conséquence [Bar05]. Dans ce contexte, l'ESD fournit de meilleurs résultats que la TMD en termes de "confiance".

Normalisation d'image

Une approche plus récente consiste à effectuer l'insertion et le décodage dans une image ayant subi un prétraitement appelé normalisation. La normalisation est une technique empruntée à la reconnaissance des formes, dont le but est d'extraire les caractéristiques d'une image invariantes aux transformations affines. Elle consiste à lui appliquer diverses transformations affines afin que ses moments géométriques satisfassent certaines conditions. En tatouage, la normalisation a pour but d'éliminer les composantes de chaque attaque géométrique affine. Le principe est donc proche de la transformation invariante : il s'agit d'un "domaine spatial invariant". Il a été d'abord proposé par Alghoniemy et Tewfik [AT00], puis étendu par Dong *et al.* [DBG+05] au tatouage multi-bit de type DS et à un plus grand nombre d'attaques affines.

Autres techniques et techniques non aveugles

Il existe d'autres méthodes de resynchronisation, telles qu'une insertion périodique de la signature ou l'insertion d'un tatouage Les transformations non affines restent également des attaques qui ne peuvent être contrées que par une utilisation de l'image originale, ou du contenu de l'image, ce qui s'éloigne du schéma additif aveugle. Notamment, Dong *et al.* [DBG+05] obtiennent de bons résultats dans un schéma non aveugle utilisant un maillage déformable qui permet de caractériser des transformations géométriques locales et aléatoires. Une fois l'attaque caractérisée, il est possible de tenter de la corriger. Notons que ce modèle peut également servir dans un schéma aveugle, en supposant le canal d'attaque connu. Le domaine de la synchronisation reste donc un problème ouvert. En particulier, les attaques géométriques peuvent être étudiées à l'aide de la théorie de l'information [TVKP05], dans le but de construire des dictionnaires de tatouage robustes à un canal avec transformation géométrique, et de calculer des capacités théoriques dans ce type de canal.

1.5.5 Exploitation des propriétés statistiques d'image en tatouage

Un algorithme de tatouage doit être adapté à tout type d'image naturelle. S'il est très difficile de trouver un modèle statistique général pour les images, on peut tirer parti des modèles existants dans les domaines transformés (cf. paragraphe 1.5.2).

Domaine spatial : pré-blanchiment au décodage

Dans le domaine spatial, une image est non-stationnaire (la luminance peut varier énormément d'une région à l'autre). On ne peut donc estimer que des propriétés statistiques locales, en approchant localement l'image par un processus quasi-stationnaire. L'idée de préfiltrer l'image tatouée avant détection afin de la blanchir a été proposée dans [DKL98]. Hernandez *et al.* proposent également d'utiliser la connaissance des moments locaux du premier et second ordre pour améliorer la détection par corrélation [HPGRN98]. Un filtre de Wiener est utilisé pour obtenir une estimation linéaire du tatouage, qui minimiserait l'erreur au sens des moindres carrés si l'image était un bruit blanc. L'image est ici considérée comme le bruit à éliminer, et le tatouage est le signal à estimer. Le filtrage de Wiener peut également être utilisé comme une attaque, si l'image est considérée comme le signal à estimer et que le tatouage comme le bruit à éliminer (cf. paragraphe 1.2.3).

Le pré-blanchiment au décodage (cf. paragraphe 1.2.5) consiste à calculer une estimation $\hat{\mathbf{x}}$ de l'image originale à partir de l'image reçue $\mathbf{z}$. Le signal préfiltré est alors $\hat{\mathbf{z}} = \mathbf{z} - \hat{\mathbf{x}}$. Pour calculer $\hat{\mathbf{x}}$, on combine le filtre de Wiener à une estimation de la moyenne locale dans une fenêtre 3x3 pour se ramener à des signaux localement stationnaires de moyenne nulle [HPGRN98] :

$$h_W(k_1,k_2) = \frac{1}{\sigma^2_{\mathbf{z}}[k_1,k_2]}\frac{1}{9}\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix} .$$

Domaine transformé : décodage optimal

Le décodeur par corrélation utilisé dans l'algorithme DS n'est optimal que pour une insertion additive et un hôte gaussien. Les modélisations des domaines transformés sont

utilisées pour optimiser la détection et le décodage (cf. paragraphe 1.2.5). La densité de probabilité $f_x(x)$ dans le domaine transformé peut être choisie parmi celles répertoriées dans la partie 1.5.2.

Dans le domaine de la DCT, $\mathbf{t_z}$ désignant la DCT par blocs de $\mathbf{z}$, (cf. paragraphe 1.5.1), le test d'hypothèses effectué est le choix de $\hat{\mathbf{m}}_l^j$ vérifiant :

$$\ln \frac{f_{\mathbf{t_z}}(\mathbf{t_z}|\mathbf{m}_l^j)}{f_{\mathbf{t_z}}(\mathbf{t_z}|\mathbf{m}_k^j)} > 0 \quad \forall k \neq l$$

Pour simplifier, notons t'_n l'échantillon n du signal $\mathbf{t_z}$ pris sous forme vectorielle. Rappelons qu'il existe un numéro de bloc k et des fréquences u, v tels que $t'_n = t^k_{\mathbf{z}}(u, v)$. On notera de plus $\hat{c}_k = c_{u,v}$ pour le différencier du code c^j_k, ainsi que $\sigma_k = \sigma_{u,v}$. On cherche donc $l \in \{1, \ldots, 2^L\}$ tel que :

$$\sum_{k=1}^{N} \frac{|t'_k + w^j_{l,k}|^{\hat{c}_k} - |t'_k - w^j_{l,k}|^{\hat{c}_k}}{\sigma_k^{\hat{c}_k}} > 0 \quad \forall k \neq l$$

Alors une statistique suffisante pour l'estimation de m^j_l est [HPG99] :

$$d^j_l = \frac{1}{P} \sum_{k \in \{(l-1)P+1, \ldots, lP\}} \frac{|t'_k + \psi_k c^j_k|^{\hat{c}_k} - |t'_k - \psi_k c^j_k|^{\hat{c}_k}}{\sigma_k^{\hat{c}_k}} \tag{1.17}$$

Dans ce chapitre, on a présenté un état de l'art du tatouage numérique. Dans le chapitre suivant, nous proposons de nouvelles techniques de tatouage par étalement de spectre, fondées sur les filtres LPTV.

Bibliographie

[AB04] A. Abrardo and M. Barni. Orthogonal dirty paper coding for informed data hiding. *Proc. SPIE*, 5306 :274–285, 2004.

[AP92] A.J. Ahumada and H.A. Peterson. Luminance-model-based DCT quantization for color image compression. *Proc. SPIE on Human Vision, Visual Proc., and Digital Display III*, 1666 :365–374, 1992.

[ARPG02] M. Alvarez-Rodríguez and F. Pérez-González. Analysis of pilot-based synchronization algorithms for watermarking of still images. *Signal Processing : Image Communication*, 17(8) :611–633, 2002.

[ASB05] O. Altun, G. Sharma, and M. Bocko. Informed watermark embedding in the fractional Fourier domain. *EUSIPCO'05*, 2005.

[ASB06] O. Altun, G. Sharma, and M. Bocko. Set theoretic quantization index modulation watermarking. *Proc. of ICASSP*, 2006.

[ASS02] I. Avcibas, B. Sankur, and K. Sayood. Statistical evaluation of image quality measures. *Journal of Electronic Imaging*, 11(2) :206–223, 2002.

[AT00] M. Alghoniemy and A.H. Tewfik. Geometric distortion correction through image normalization. *ICME Multimedia Expo*, 2000.

[Aut02] F. Autrusseau. *Tatouage d'images fondé sur la modélisation du système visuel hmain et sur la transformation Mojette*. PhD thesis, Ecole polytechnique de l'Université de Nantes, 2002.

[Bar03] M. Barni. What is the future for watermarking ? (part 1). *IEEE Signal Processing Magazine*, 20(5) :55–60, 2003.

[Bar05] M. Barni. Effectiveness of exhaustive search and template matching against watermark desynchronization. *IEEE Signal Proc. Letters*, 12(2) :158–161, 2005.

[Bas05] P. Bas. A quantization watermarking technique robust to linear and non-linear valumetric distortions using a fractal set of quantizers. *Information Hiding Workshop, Proc.*, pages 83–93, 2005.

[BB04] M. Barni and F. Bartolini. Data hiding for fighting piracy. *Signal Processing Magazine, IEEE*, 21(2) :28–39, 2004.

[BBF03] M. Barni, F. Bartolini, and T. Furon. A general framework for robust watermarking security. *Signal Processing*, 83(10) :2069–2084, 2003.

[BBP04] F. Bartolini, M. Barni, and A. Piva. Performance analysis of st-dm watermarking in presence of nonadditive attacks. *IEEE Trans. on Signal Processing*, 52(10) :2965–2974, 2004.

[BBPR98] M. Barni, F. Bartolini, A. Piva, and F. Rigacci. Statistical modelling of full frame DCT coefficients. *Proc. of EUSIPCO'98*, pages 1513–1516, 1998.

[BBRP00] M. Barni, F. Bartolini, A. De Rosa, and A. Piva. Capacity of full frame dct image watermarks. *IEEE Trans. on Image Processing*, 9(8) :1430–1435, 2000.

[BC03] D. Barba and P. Le Callet. A robust quality metric for color image quality assessment. *Proc. of IEEE ICIP*, 1 :437–440, 2003.

[BC06a] P. Bas and F. Cayre. Achieving Subspace or Key Security for WOA using Natural or Circular Watermarking. *ACM Multimedia and Security*, 2006.

[BC06b] P. Bas and F. Cayre. Natural watermarking : a secure spread spectrum technique for WOA. *Information Hiding Workshop, Proc.*, 2006.

[BCD98] P. Bas, J-M. Chassery, and F. Davoine. Self-similarity based image watermarking. *Proc. of EUSIPCO*, 4 :2277–2280, 1998.

[BCD99] P. Bas, J-M. Chassery, and F. Davoine. A geometrical and frequential watermarking scheme using similarities. *Proc of SPIE Elec. Imaging, Security and Wat. of Multimedia Content I*, pages 264–272, 1999.

[BCM00] P. Bas, J-M. Chassery, and B. Macq. Robust watermarking based on the warping of pre-defined triangular patterns. *Proc of SPIE Elec. Imaging, Security and Wat. of Multimedia Content II*, 18 :99–109, 2000.

[BDBT06a] J.-P. Boyer, P. Duhamel, and J. Blanc-Talon. Asymptotically Optimal Scalar Quantizers for QIM Watermark Detection. *Proc. of ICME*, 2006.

[BDBT06b] J.-P. Boyer, P. Duhamel, and J. Blanc-Talon. Performance analysis of scalar DC-QIM for watermark detection. *Proc. of ICASSP*, 2006.

[BGML96] W. Bender, D. Gruhl, N. Morimoto, and A. Lu. Techniques for data hiding. *IBM Systems Journal*, 35 :313–336, 1996.

[BKMS01] N. Boulgouris, I. Kompatsiaris, V. Mezaris, and M. Strintzis. Content-based watermarking for indexing using robust segmentation. *Proc. Workshop on Image Analysis For Multimedia Interactive Services (WIAMIS)*, 2001.

[BKZ98] S. Burgett, E. Koch, and J. Zhao. Copyright labelling of digitized image data. *IEEE Commun. Mag.*, 36 :94–100, 1998.

[Bla87] R. Blahut. *Principes and practice of information theory*. Addisson-Wesley, 1987.

[BM01] P. Bas and B. Macq. Tatouage d'objets vidéos résistant aux manipulations. *Traitement du Signal*, 18(18) :249–257, 2001.

[Bru03] H. Brunk. Host-aware spread spectrum watermark embedding techniques. *Proc. of SPIE Security and Watermarking of Multimedia Contents V*, 5020 :699–707, 2003.

[BS95] D. Boneh and J. Shaw. Collusion-secure fingerprinting for digital data. *Proc. of CRYPTO*, 1995.

[BSNO99] G. Brisbane, R. Safavi-Naini, and P. Ogunbona. Region-based watermarking for images. *Information Security Workshop (ISW)*, pages 154–166, 1999.

[CCB03] M. Carnec, P. Le Callet, and D. Barba. An image quality assessment method based on perception of structural information. *Proc. of IEEE ICIP*, 3 :185–188, 2003.

[CFF05] F. Cayre, C. Fontaine, and T. Furon. Watermarking security : Theory and practice. *IEEE Trans. on Signal Processing, Special Issue on Content Protection*, 53(10) :3976–3975, 2005.

[Che] CheckMark. http ://watermarking.unige.ch/Checkmark/.

[CKLS96] I.J. Cox, J. Killian, T. Leighton, and T. Shamoon. Secure Spread Spectrum Watermarking for Images, Audio, and Video. *IEEE ICIP'96*, III :243–246, 1996.

[CKLS97] I.J. Cox, J. Kilian, F.T. Leighton, and T. Shamoon. Secure spread spectrum watermarking for multimedia. *Image Processing, IEEE Trans. on*, 6(12) :1673–1687, 1997.

[CL04] C. Crampes and J. Larrieu. Aspects économiques et juridiques de la propriété intellectuelle. *Formation CIES, Université des Sciences Sociales de Toulouse*, 2004.

[CMB02] I.J. Cox, M.L. Miller, and J.A. Bloom. *Digital Watermarking*. Morgan Kaufmann Publisher, Inc., San Francisco, 2002.

[CMM99] I.J. Cox, M.L. Miller, and A.L. McKellips. Watermarking as communications with side information. *Proceedings of the IEEE, Special Issue on Identification and Protection of Multimedia Information*, 87(7) :1127–1141, 1999.

[Cos83] M.H.M. Costa. Writing on dirty paper. *IEEE Trans. on Information Theory*, 29(3) :439–441, 1983.

[CP95] J.-P. Cocquerez and S. Philipp. *Analyse d'images : filtrage et segmentation*. Masson, 1995.

[CPFPG05a] P. Comesaña, L. Pérez-Freire, and F. Pérez-González. Fundamentals of data-hiding security and their application to spread-spectrum analysis. *Information Hiding Workshop*, 3727 :122–136, 2005.

[CPFPG05b] P. Comesaña, L. Pérez-Freire, and F. Pérez-González. An information-theoretic framework for assessing security in practical watermarking and data hiding scenarios. *6th International Workshop on Image Analysis for Multimedia Interactive Services*, 2005.

[CPFPG06] P. Comesaña, L. Pérez-Freire, and F. Pérez-González. The blind newton senstivity attack. *SPIE*, 2006.

[CPG06] P. Comesaña and F. Pérez-González. The impact of the cropping attack on scalar STDM data hiding. *IEEE Signal Processing Letters*, 2006.

[CPR99] J. Chou, S.S. Pradhan, and K. Ramchandran. On the duality between distributed source coding and data hiding. *Proc. Asilomar Conference on Signals, Systems, and Computers*, 2 :2061–2064, 1999.

[CPR01] J. Chou, S.S. Pradhan, and K. Ramchandran. Turbo coded trellis-based constructions for data embedding : channel coding with side information. *Signals, Systems and Computers, 2001*, 1 :305 – 309, 2001.

[Cra98] S. Craver. Resolving rightful ownerships with invisible watermarking techniques : limitations, attacks, and implications. *Selected Areas in Communications, IEEE Journal on*, 16(4) :573–586, 1998.

[CSST01] M.U. Celik, E. Saber, G. Sharma, and A.M. Tekalp. Analysis of feature-based geometry invariant watermarking. *Proc. SPIE on Security and Watermarking of Multimedia Contents III*, 4314 :261–268, 2001.

[CT03] N. Cvejic and T.Seppänen. Increasing robustness of an improved spread spectrum audio watermarking method using attack characterization. *Proc. International Workshop on Digital Watermarking*, pages 467–473, 2003.

[CW01] B. Chen and G.W. Wornell. Quantization index modulation : A class of provably good methods for digital watermarking and information embedding. *IEEE Trans. on Information Theory*, pages 1423–1443, 2001.

[Dal94] S. Daly. A visual model for optimizing the design of image processing algorithms. *Proc. of ICIP*, pages 16–20, 1994.

[DBG+05] P. Dong, J.G. Brankov, N.P. Galatsanos, Y. Yang, and F. Davoine. Digital watermarking robust to geometric distortions. *IEEE Trans. on Image Proc.*, 14(12) :2140–2150, 2005.

[dCTG02] L. de Campos Teixeira Gomes. *Tatouage de signaux audio*. PhD thesis, Université René Descartes - Paris V, 2002.

[DD05a] G.J. Doërr and J.-L. Dugelay. Collusion number in video watermarking. *Security, Steganography, and Watermarking of Multimedia Content VII, Proceedings of SPIE*, 5681, 2005.

[DD05b] G.J. Doërr and J.-L. Dugelay. How to combat block replacement attacks ? *Information Hiding Workshop*, pages 137–151, 2005.

[DDML02] J.F. Delaigle, C. Devleeschouwer, B. Macq, and L. Langendijk. Human visual system features enabling watermarking. *Multimedia and Expo,*

2002. ICME '02. Proceedings. 2002 IEEE International Conference on, 2 :26–29, 2002.

[DFHS03] J. Delhumeau, T. Furon, N. Hurley, and G. Silvestre. Improved polynomial detectors for side-informed watermarking. *Proc. SPIE*, 2003.

[DKL98] G. Depovere, T. Kalker, and J.-P. Linnartz. Improved watermark detection using filtering before correlation. *Proc. if IEEE ICIP*, 1 :430–434, 1998.

[DM03] M.N. Desai and R.S. Mangoubi. Robust gaussian and non-gaussian matched subspace detection. *Signal Processing, IEEE Transactions on*, 51(12) :3115–3127, 2003.

[DM06] D. Delannay and B. Macq. Watermarking relying on cover signal content to hide synchronization marks. *IEEE Transactions on Information Forensics and Security*, 1 :87–101, 2006.

[DP04] F. Davoine and S. Pateux. *Tatouage de documents audiovisuels numériques*. Hermes Science, 2004.

[DRRD06] J.L. Dugelay, S. Roche, C. Rey, and G. Doerr. Still-image watermarking robust to local geometric distortions. *IEEE Trans. on Image Processing*, 15(9), 2006.

[EBG02] J.J. Eggers, R. Bauml, and B. Girod. Estimation of amplitude modifications before scs watermark insertion. *Proc. SPIE*, pages 387–398, 2002.

[EBTG03] J.J. Eggers, R. Bäuml, R. Tzschoppe, and B. Girod. Scalar Costa Scheme for Information Embedding. *IEEE Trans. on Signal Processing*, 51(4) :1003–1019, 2003.

[EG00] J.J. Eggers and B. Girod. Quantization Watermarking. *SPIE Security and Watermarking of Multimedia Contents II*, 2000.

[EG01] J.J. Eggers and B. Girod. Quantization effects on digital watermarks. *EURASIP Signal Processing*, 81(2) :239–263, 2001.

[F. 93] F. Müller. Distribution shape of two-dimensional DCT coefficients of natural images. *Electronic Letters*, 29(22) :1953–1954, 1993.

[FD03] T. Furon and P. Duhamel. An asymmetric watermarking method. *IEEE. Trans. on Signal Proc.*, 51(4) :981–995, 2003.

[FG01] J. Fridrich and M. Goljan. Invertible authentication watermark for jpeg images. *ITCC*, 2001.

[FG02] J. Fridrich and M. Goljan. Lossless data embedding - new paradigm in digital watermarking. *EURASIP Journal on Applied Signal Processing*, 2 :185–196, 2002.

[FKK04] C. Fei, D. Kundur, and R.H. Kwong. Analysis and Design of Watermarking Algorithms for Improved Resistance to Compression. *IEEE Trans. on Image Processing*, 13(2) :126–144, 2004.

[FMSH02] T. Furon, B. Macq, G. Silvestre, and N. Hurley. JANIS : Just Another N-order side-Informed watermarking Scheme. *Proc. of Int. Conf. on Image Processing ICIP'02*, 2002.

[FS02] V. Fotopoulos and A.N. Skodras. JPEG2000 Parameters Against Watermarking. *Proc. 14th Int. Conf. on Digital Signal Processing (DSP2002)*, 2 :713–716, 2002.

[FSL04] Y. Fu, R. Shen, and H. Lu. Optimal watermark detection based on support vector machines. *Int. Symposium on Neural Networks (ISNN)*, pages 552–557, 2004.

[FTB04] R. F. H. Fischer, R. Tzschoppe, and R. Bäuml. Lattice Costa Schemes using Subspace Projection for Digital Watermarking. *European Trans. on Telecommunications (ETT)*, 15(4) :351–362, 2004.

[Fur02] T. Furon. *Use of watermarking techniques for copy protection*. PhD thesis, Ecole Nationale Supérieure des Télécommunications, 2002.

[GDV+97] F. Goffin, J.-F. Delaigle, C. De Vleeschouwer, B. Macq, and J.-J. Quisquater. A low cost perceptive digital picture watermarking method. *SPIE Electronic Imaging*, 3022(28) :264–277, 1997.

[GFVFVRS01] J.A. García, J. Fdez-Valdivia, X.R. Fdez-Vidal, and R. Rodriguez-Sánchez. Information theoretic measure for visual target distinctness. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 23(4) :362–383, 2001.

[GLB06] F. Guerrini, R. Leonardi, and M. Barni. Image watermarking robust against non-linear value-metric scaling based on higher order statistics. *Proc. of ICASSP*, 2006.

[GM02] L. Guillemot and J.-M. Moureaux. Bit-rate adapted watermarking algorithm for compressed images. *Proc. of ICME*, pages 545–548, 2002.

[GM04] L. Guillemot and J.-M. Moureaux. Hybrid transmission, compression and data hiding : quantisation index modulation as source coding strategy. *Electronics Letters*, 40(17) :1053–1055, 2004.

[GP80] S.I. Gel'fand and M.S. Pinsker. Coding for channel with random parameters. *Problems of Control and Information Theory*, 9(1) :19–31, 1980.

[GP03] G. Le Guelvouit and S. Pateux. Wide spread spectrum watermarking with side information and interference cancellation. *Proc. SPIE*, 2003.

[Gre02] M.L. Green. Statistics of images, the TV algorithm of Rudin-Osher-Fatemi for image denoising and an improved denoising algorithm. *CAM reports, Univ. California, Los Angeles [Online] : http ://www.math.ucla.edu/applied/cam/index.html*, 2002.

[Gue03] G. Le Guelvouit. *Tatouage robuste par étalement de spectre avec prise en compte de l'information adjacente*. PhD thesis, Rennes, INSA, 2003.

[HAPG00] J.R. Hernández, M. Amado, and F. Pérez-González. DCT-Domain Watermarking Techniques for Still images : Detector Performance analysis and a New Structure. *IEEE Trans. on Image Processing*, 9(1) :55–68, 2000.

[Har99] F. Hartung. *Digital Watermarking and Fingerprinting of Uncompressed and Compressed Video*. Shaker Verlag, Aachen, Germ., 99.

[Her02] C. Herley. Why watermarking is nonsense. *Signal Processing Magazine, IEEE*, 19(5) :10–11, 2002.

[HK99] F. Hartung and M. Kutter. Multimedia watermarking techniques. *Proc. of the IEEE*, 87(7) :1079–1107, 1999.

[HM99] J. Huang and D. Mumford. Statistics of natural images and models. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'99)*, 1 :541–547, 1999.

[HPG99] J.R. Hernández and F. Pérez-González. Statistical analysis of watermarking schemes for copyright protection of images. *IEEE Proc., Special Issue on Identification and Protection of Multimedia Information*, 87(7) :1142–1166, 1999.

[HPGRN98] J.R. Hernández, F. Pérez-González, J.M. Rodriguez, and G. Nieto. Performance analysis of a 2-D-multipulse amplitude modulation scheme for data hiding and watermarking of still images. *Selected Areas in Communications, IEEE Journal on*, 16(4) :510–524, 1998.

[HW00] C.-H. Huang and J.-L. Wu. A watermark optimization technique based on genetic algorithms. *Proc. SPIE*, 3971 :516–523, 2000.

[Jai89] A.K. Jain. *Fundamentals of Digital Image Processing*. Prentice-Hall International, 1989.

[JDB96] J.J.K. Ó Ruanaith, W. Dowling, and F. Boland. Phase watermarking of digital images. *Proc. of ICIP*, 3 :239–242, 1996.

[JF95] R.L. Joshi and T.R. Fischer. Comparison of Generalized Gaussian and Laplacian Modelling in DCT Image Coding. *IEEE Signal Processing Letters*, 2(5) :81–82, 1995.

[JHM01] M. Kutter J.R. Hernandez Martin. Information retrieval in digital watermarking. *IEEE Communications Magazine*, August :110–116, 2001.

[JKB95] N.L. Johnson, S. Kotz, and N. Balakrishnan. *Continuous Multivariate Distributions, Vol. 2*. Wiley-Interscience, 2nd edition, 1995.

[JP98] J.J.K. Ó Ruanaith and T. Pun. Rotation, scale and translation invariant spread spectrum digital image watermarking. *Signal Proc.*, 66(3) :303–317, 1998.

[KA05] P. Kumsawat and K. Attakitmongcol. A new approach for optimization in image watermarking using genetic algorithms. *IEEE Trans. on Signal Proc.*, 53(12), 2005.

[Kat05] S. Katzenbeisser. Computational Security Models for Digital Watermarks. *Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS)*, 2005.

[KH98] D. Kundur and D. Hatzinakos. Digital Watermarking Using Multiresolution Wavelet Decomposition. *IEEE ICASSP*, 5 :2659–2662, 1998.

[KJ99] T. Kalker and A. Janssen. Analysis of SPOMF detection. *Proc. of IEEE conference on ICIP*, 1 :316–319, 1999.

[KL02] M. Koppen and Xiufen Liu. Content-based watermarking using image texture. *6th International Conference on Signal Processing*, 2 :1576 – 1579, 2002.

[Koz02] A. Koz. *Digital Watermarking based on the Human Visual System*. PhD thesis, Middle East Technical University, 2002.

[KP03a] D. Kirovski and F. Petitcolas. Blind pattern matching attack on watermarking systems. *IEEE Trans. on Signal Processing*, 51(4) :1045–1053, 2003.

[KP03b] D. Kirovski and F.A.P. Petitcolas. Blind pattern matching attack on watermarking systems. *IEEE Transactions on signal processing*, 51(4) :1045–1053, 2003.

[LL05] C.-H. Lee and H.-K. Lee. Improved autocorrelation function based watermarking with side information. *Journal of Electronic Imaging*, 14(1), 2005.

[LLH05] H. Liu, J. Lin, and J. Huang. Image authentication using content based watermark. *IEEE Int. Symp. on Circuits and Systems (ISCAS)*, 4 :4014–4017, 2005.

[LMH01] A.B. Lee, D. Mumford, and J. Huang. Occlusion Models for Natural Images : A Statistical Study of a Scale-Invariant Dead Leaves Model. *Int. Journal of Computer Vision*, 41(1-2) :35–59, 2001.

[LS04] R.L. Lagendijk and I.D. Shterev. Estimation of attacker's scale and noise variance for qim-dc watermark embedding. *Proc. of ICIP*, 2004.

[LSL00] G.C. Langelaar, I. Setyawan, and R.L. Lagendijk. Watermarking Digital Image and Video Data : A State-of-the-Art Overview. *IEEE Signal Processing Magazine*, 17(5) :20–46, 2000.

[LWB+01] C.Y. Lin, M. Wu, J.A. Bloom, I.J. Cox, M. Miller, and Y.M. Liu. Rotation, scale and translation resilient public watermarking for images. *IEEE Trans. on Image Proc.*, 9(6) :767–782, 2001.

[MACC07] E. Marini, F. Autrusseau, P. Le Callet, and P. Campisi. Evaluation of standard watermarking techniques. *SPIE Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents IX*, 2007.

[Mai02] H. Maitre. *Le traitement des images*. Hermès, 2002.

[Mal89] S.G. Mallat. A theory for multiresolution signal decomposition : the wavelet representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 11(7) :674–693, 1989.

[MCB00] M.L. Miller, I.J. Cox, and J.A. Bloom. Informed embedding : Exploiting image and detector information during watermark insertion. *IEEE Int. Conf. on Image Processing - ICIP*, 3 :1–4, 2000.

[MDC04] M.L. Miller, G.J. Doërr, and I.J. Cox. Applying informed coding and embedding to design a robust high-capacity watermark. *IEEE Trans. on image processing*, 13(6) :792–807, 2004.

[Mee01] P. Meerwald. Quantization watermarking in the JPEG2000 coding pipeline. *Proc. of IFIP-CMS*, pages 69–79, 2001.

[Mer05] N. Merhav. On joint coding for watermarking and encryption. *Information Hiding Workshop*, pages 1–10, 2005.

[MF03] H.S. Malvar and D.A.F. Florêncio. Improved spread spectrum : a new modulation technique for robust watermarking. *IEEE Trans. on Signal Processing*, 51(4) :898–905, 2003.

[MK05] P. Moulin and R. Koetter. Data-hiding codes. *Proc. of the IEEE*, 93(12) :2083–2127, 2005.

[MM06a] B. Michiels and B. Macq. Benchmarking image watermarking algorithms with openwatermark. *Proc. of EUSIPCO*, 2006.

[MM06b] V. Monga and K. Mihcak. Robust image hashing via non-negative matrix factorizations. *Proc. of ICASSP*, 2006.

[MO98] M.J.J.J.B. Maes and C.W.A.M Overveld. Digital watermarking by geometric warping. *Proc of ICIP*, 2 :424–426, 1998.

[MO03] P. Moulin and J.A. O'Sullivan. Information-theoretic analysis of information hiding. *IEEE Transactions on Information Theory*, 49(3) :563–593, 2003.

[Mou02] P. Moulin. Information-hiding games. *Int. Workshop on Digital Watermarking*, 2002.

[Mou03] P. Moulin. What is the future for watermarking ? (part 2). *IEEE Signal Processing Magazine*, 20(6) :51–57, 2003.

[MSB02] J. Mayer, A.V. Silvério, and J.C.M. Bermudez. On the design of pattern sequences for spread spectrum image watermarking. *Proc. International Telecommunications Symposium (ITS-2002), Natal, RN, Brazil*, 2002.

[MSS05] A. Martin, G. Sapiro, and G. Seroussi. Is image steganography natural ? *IEEE Trans. on Image Processing*, 14(12) :2040–2050, 2005.

[MT02] M. Mansour and A. Tewfik. Secure detection of public watermarks with fractal decision boundaries. *IEEE EUSIPCO'02, Proc.*, 1 :295–298, 2002.

[MVK02] M.K. Mihcak, R. Venkatesan, and M. Kesal. Cryptanalysis of discrete-sequence spread spectrum watermarks. *Proc. of the 5th International Information Hiding Workshop (IH 2002)*, 2002.

[NP98] N. Nikolaidis and I. Pitas. Robust image watermarking in the spatial domain. *Signal Processing*, 66(3) :385–403, 1998.

[OKS04] J. Oostven, T. Kalker, and M. Staring. Adaptive quantization watermarking. *Proc. of SPIE*, 5306 :296–303, 2004.

[Oli98] J.-L. Olivès. *Optimisation Globale d'un système imageur à l'aide de critères de qualité visuelle*. PhD thesis, ENAC, 1998.

[PBBC97] A. Piva, M. Barni, F. Bartolini, and V. Cappellini. DCT-based watermark recovering without resorting to the uncorrupted original image. *ICIP'97, Proc.*, 1 :520–523, 1997.

[Pet00] F.A.P. Petitcolas. Watermarking schemes evaluation. *IEEE Signal Processing*, 17(5) :58–64, 2000.

[PFCPG05a] L. Pérez-Freire, P. Comesaña, and F. Pérez-González. Detection in quantization-based watermarking : performance and security issues. *SPIE*, 2005.

[PFCPG05b] L. Pérez-Freire, P. Comesaña, and F. Pérez-González. Information-theoretic analysis of security in side-informed data hiding. *Information Hiding Workshop, Proc.*, pages 107–121, 2005.

[PFCTPPG06] L. Pérez-Freire, P. Comesaña, J.R. Troncoso-Pastoriza, and F. Pérez-González. Watermarking security : a survey. *LNCS Transactions on Data Hiding and Multimedia Security. To appear.*, 2006.

[PFPG05] L. Pérez-Freire and F. Pérez-González. Spread-spectrum vs. quantization-based data hiding : misconceptions and implications. *SPIE Security, Steganography, and Watermarking of Multimedia Contents*, VII, 2005.

[PFPGV06] L. Pérez-Freire, F. Pérez-González, and S. Voloshynovskiy. An accurate analysis of scalar quantization-based data-hiding. *IEEE Trans. on Information Forensics and Security*, 1(1) :80–86, 2006.

[PGB02] F. Pérez-González and F. Balado. Improving data hiding performance by using quantization in a projected domain. *Proc. of the IEEE International Conference on Multimedia and Expo (ICME), Lausanne, Switzerland*, 2002.

[PGBM03] F. Pérez-González, F. Balado, and J.R. Hernández Martin. Performance analysis of existing and new methods for data hiding with known-host information in additive channels. *IEEE Trans. on Signal Processing*, 51(4) :960–980, 2003.

[PGCB03] F. Pérez-González, P. Comesaña, and F. Balado. Dither-modulation data hiding with distortion-compensation : exact performance analysis and an improved detector for jpeg attack. *IEEE International Conference on Image Processing (ICIP)*, 2003.

[PGMBA05] F. Pérez-González, C. Mosquera, M. Barni, and A. Abrardo. Rational dither modulation : A high-rate data-hiding method invariant to gain attacks. *IEEE Trans. on Signal Processing*, 10(2) :3960–3975, 2005.

[PJ96] J. Puate and F. Jordan. Using fractal compression scheme to embed a signature into an image. *Proc. of the SPIE Photonics East'96 Symposium*, pages 108–118, 1996.

[PP03] S. Pereira and T. Pun. Fast robust template matching for affine resistant image watermarking. *IEEE Transactions on signal proc.*, 51(4) :1045–1053, 2003.

[PTB+05] T. Pasquier, P. Treguer, C. Bareille, S. Bois, S. Martin, and A. Couillaud. Rencontre avec le droit d'auteur. *Livre d'accompagnement de l'Exposition, Espace Mendès France de Poitiers*, 2005.

[PVM+01] S. Pereira, S. Voloshynovskiy, M. Madueño, S. Marchand-Maillet, and T. Pun. Second generation benchmarking and application oriented evaluation. *Information Hiding Workshop III*, 2001.

[PVP01] S. Pereira, S. Voloshynoskiy, and T. Pun. Optimal transform domain watermark embedding via linear programming. *Signal Processing*, 81(6) :1251–1260, 2001.

[RG83] R.C. Reiniger and J.D. Gibson. Distribution of the two-dimensional DCT coefficients for image. *IEEE Trans. on Communications*, 31 :835–839, 1983.

[SC96] J.R. Smith and B.O. Comiskey. Modulation and information hiding in images. *Information Hiding Workshop*, pages 207–226, 1996.

[SEG01] J. Su, J. Eggers, and B. Girod. Analysis of digital watermarks subjected to optimum linear filtering and additive noise. *Signal Processing*, 81 :1141–1175, 2001.

[SHWP04] C.-S. Shieh, H.-C. Huang, F.-H. Wang, and J.-S. Pan. Genetic watermarking based on transform-domain techniques. *Pattern Recognition*, 37(3) :555–565, 2004.

[SLSZ03] A. Srivastava, A.B. Lee, E.P. Simoncelli, and S.-C. Zhu. On advances in statistical modelling of natural imges. *Journal of Mathematical Imaging and Vision*, 18(1) :17–33, 2003.

[SMCM05] K. Sullivan, U. Madhow, S. Chandrasekaran, and B. S. Manjunath. Steganalysis of spread spectrum data hiding exploiting cover memory. *Proc. SPIE*, pages 38–46, 2005.

[SNZ+04] Y.Q. Shi, Z. Ni, D. Zou, C. Liang, and G. Xuan. Lossless data hiding : fundamentals, algorithms and applications. *Circuits and Systems, ISCAS '04. Proceedings of the 2004 International Symposium on*, 2(2) :33–36, 2004.

[Sti] Stirmark. http ://www.petitcolas.net/fabien/watermarking/stirmark/.

[SWA94] J.A. Solomon, A.B. Watson, and A.J. Ahumada. Visibility of DCT basis functions : Effects of contrast masking. *Proc. Data Compression Conf.*, pages 361–370, 1994.

[TRvS+93] A.Z. Tirkel, G.A. Rankin, R.M. van Schyndel, W.J. Ho, N.R.A. Mee, and C.F. Osborne. Electronic water mark. *DICTA'93 - Digital Image Computing, Technology and Applications*, pages 666–673, 1993.

[TVK+05] E. Topak, S. Voloshynovskiy, O. Koval, J.E. Vila-Forcén, and T. Pun. On security of geometrically-robust data-hiding. *WIAMIS, 6th International Workshop on Image Analysis for Multimedia Interactive Services*, 2005.

[TVKP05] E. Topak, S. Voloshynovskiy, O. Koval, and T. Pun. Achievable rate analysis of geometrically robust data-hiding codes in asymptotic setups. *Proc. of EUSIPCO'05*, 2005.

[vdBLF96] C.J. van den Branden Lambrecht and J.E. Farrel. Perceptual quality metric for digitally coded color images. *Proc. of EUSIPCO*, pages 1175–1178, 1996.

[VDP01] S. Voloshynovskiy, F. Deguillaume, and T. Pun. Multibit digital watermarking robust against local nonlinear geometrical distortions. *Porc. of ICIP*, pages 999–1002, 2001.

[VHBP99] S. Voloshynovskiy, A. Herrigel, N. Baumgartner, and T. Pun. A stochastic approach to content adaptive digital image watermarking. *International Workshop on Information Hiding*, pages 212–236, 1999.

[VPP+01] S. Voloshynovskiy, S. Pereira, T. Pun, J.J. Eggers, and J.K. Su. Attacks on digital watermarks : classification, estimation based attacks, and benchmarks. *Communications Magazine, IEEE*, 39(8) :118–126, 2001.

[Wat93] A.B. Watson. Visually Optimal DCT Quantization Matrices for Individual Images. *Data Compression Conference*, pages 178–187, 1993.

[WBL02] Z. Wang, A.C. Bovik, and Ligang Lu. Why is image quality assessment so difficult ? *ICASSP*, 4 :3313–3316, 2002.

[WBSS04] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment : From error visibility to structural similarity. *IEEE Trans. on Image Proc.*, 13(4) :600–612, 2004.

[WCA04] P.H.W. Wong, A. Chang, and O.C. Au. A Sequential Multiple Watermarks Embedding Technique. *IEEE ICASSP'04. Proc.*, 3 :393–396, 2004.

[Web06] Webzine L'internaute. Le projet de loi sur les droits d'auteurs adopté par l'assemblée : ce qu'il faut retenir. *http :/www.linternaute.com, mars 2006*, 2006.

[WL04] S.-H. Wang and Y.-P. Lin. Wavelet tree quantization for copyright protection watermarking. *IEEE Trans. on Image Proc.*, 13(2), 2004.

[WLVW00] R.A. Wannamaker, S.P. Lipshitz, J. Vanderkooy, and J.N. Wright. A Theory of Nonsubtractive Dither. *IEEE Trans. on Signal Proc.*, 48(2) :499–516, 2000.

[WYA03] P.H.W. Wong, G.Y.M. Yeung, and O.C. Au. Capacity for JPEG2000-to-JPEG2000 images watermarking. *Proc. of ICME*, 2 :485–488, 2003.

[WYSV97] A.B. Watson, G.Y. Yang, J.A. Solomon, and J. Villasenor. Visibility of wavelet quantization noise. *Image Processing, IEEE Transactions on*, 6(8) :1164–1175, 1997.

[XLP04] Yongqing Xin, S. Liao, and M. Pawlak. A multibit geometrically robust image watermark based on zernike moments. *Proc. of ICPR (Pattern Recognition)*, 4 :861–864, 2004.

[YM97] M.M. Yeung and F. Mintzer. An invisible watermarking technique for image verification. *Proc. of ICIP*, 2 :26–29, 1997.

[YTL01] P.-T. Yu, H.-H. Tsai, and J.-S. Lin. Digital watermarking based on neural networks for color images. *Signal Processing Magazine*, 81(3) :663–671, 2001.

[ZD05] A. Zaidi and P. Duhamel. Joint source-channel coding for lattice watermarking. *European Signal Processing Conference, EUSIPCO*, 2005.

[ZP06] A. Zaidi and P. Piantanida. Mac aware coding strategy for multiple user information embedding. *Proc. of ICASSP*, 2006.

[ZPD05] A. Zaidi, P. Piantanida, and P. Duhamel. Scalar scheme for multiple user information embedding. *Proc. of ICASSP*, 2, 2005.

[ZSS02] R. Zamir and U. Erez S. Shamai. Nested linear/lattice codes for structured multiterminal binning. *IEEE Transactions on Information Theory*, 48(6) :1250–1276, 2002.

Chapitre 2

Tatouage par étalement de spectre fondé sur les filtres LPTV

Sommaire

Ce chapitre présente des techniques de tatouage par étalement de spectre. Les filtres Linéaires Variant Périodiquement dans le Temps (LPTV) y sont utilisés comme alternative à l'étalement classique utilisant la modulation par une séquence pseudo-aléatoire. Dans un premier temps, nous étudions le cas particulier simple des Changements d'Horloge Périodiques (PCC). Nous montrons ensuite que le choix du parcours spatial d'une image, pour passer d'un signal bidimensionnel à un signal unidimensionnel, a un impact significatif sur les performances. Cette amélioration profite également au tatouage par filtres LPTV. De plus, dans ce cadre plus général, il est possible de construire des filtres adaptés au cadre du tatouage numérique qui permettent, simultanément à l'étalement, d'éliminer le bruit dû à l'image hôte ou encore d'opérer un masquage spectral. La chaîne de tatouage utilisant les filtres LPTV est complétée par des extensions utilisant les propriétés statistiques de l'image au décodage. La combinaison des filtres LPTV avec les techniques de tatouage informé impliquant un étalement (étalement de spectre amélioré et techniques quantificatives) est ensuite étudiée. Enfin, nous évaluons le niveau de sécurité théorique des techniques envisagées et proposons un algorithme pratique d'attaque sur la sécurité des PCC.

Ce chapitre fera parfois référence à l'annexe 2.3.1, qui rassemble notamment des variantes, ainsi que la plupart des simulations comparant la robustesse des algorithmes proposés et des algorithmes classiques.

2.1 Tatouage par Changements d'Horloge Périodiques (PCC)

2.1.1 Présentation des changements d'horloge périodiques

Définition et propriétés

Soit f une fonction T-périodique de n. Dans un cadre stochastique, si $U = \{u_n, n \in \mathbf{Z}\}$ est un processus stationnaire, on appelle Changement d'Horloge Périodique (PCC) [Lac96] le nouveau processus Y :

$$v_n = u_{n-f(n)}, \qquad \text{où} \quad f(n) = f(n+T) \ .$$

Pour une suite numérique, un PCC équivaut donc à un déplacement périodique, mais pas nécessairement aléatoire ni inversible, des échantillons. Les changements d'horloge servent à modéliser de nombreux phénomènes physiques, le principe étant toujours de prendre en compte des retards différents d'un même processus : propagation des ondes acoustiques, rétrodiffusion d'ondes radar sur les arbres, propagation d'un signal dans la bande de fréquences HF... Les changements d'horloge périodiques modélisent plus particulièrement certains effets de diffraction acousto-optique, ou encore le comportement d'un gaz soumis à un champ magnétique intense [Lac00]. Les conditions d'inversibilité des PCC sont étudiées dans [Cha04]. La famille des PCC linéaires définis par $f(n) = -k\lfloor n \rfloor$, où $\lfloor n \rfloor$ est le reste de la division euclidienne de n par T ($n = \lceil n \rceil T + \lfloor n \rfloor$), est un exemple de PCC inversible.

On étudiera dans la suite un cas particulier de PCC fondé sur les permutations aléatoires et qui sera utile pour les télécommunications, grâce à ses propriétés d'étalement

de spectre. Soit f une permutation aléatoire T-périodique définie par :

$$f(n) = \lfloor n \rfloor - q(\lfloor n \rfloor)$$

où q est une permutation de $(0, 1, 2, \ldots, T-1)$. Comme q est une permutation, q est inversible et le PCC l'est également. La fonction périodique inverse de f est f^{-1} définie par :

$$f^{-1}(n) = \lfloor n \rfloor - q^{-1}(\lfloor n \rfloor) \ .$$

Ainsi, $v_n = u_{\lceil n \rceil T + q(\lfloor n \rfloor)}$. La nouvelle suite Y est de moyenne nulle et cyclostationnaire [Gar94] (cf. 2.3.1). On peut stationnariser Y en considérant le processus $v_{n+\phi}$, où ϕ est une variable aléatoire uniformément distribuée sur $\{0, 1, \ldots, T-1\}$. De plus, pour T suffisamment grand, le spectre de V s'approche de celui d'un bruit blanc. Cette propriété est démontrée dans l'annexe B.1.1, qui reprend les calculs de [LR02].

Utilisation pour les télécommunications

Les communications multi-utilisateurs utilisant les PCC transmettent une permutation aléatoire f_j de même période T de chaque message. L'application successive de deux PCC quelconques $f_i \circ f_j$ est un PCC et étale le spectre. Seul le PCC inverse f_j^{-1} permet de retrouver le signal d'entrée. Ces propriétés d'étalement et d'orthogonalité sont démontrées pour $T \to \infty$. Le plus souvent, ce sont les permutations aléatoires périodiques qui sont utilisées pour leurs propriétés de blanchiment du spectre. Outre les PCC linéaires déjà évoqués, le cas particulier de l'entrelaceur ligne/colonne, courant en télécommunications et modélisable par un PCC, est également étudié [Cha04]. Ces deux dernières familles de PCC ne présentent cependant pas les garanties de sécurité nécessaires à une application au tatouage numérique. D'autre part, il a été montré que l'ensemble des PCC est confondu avec l'ensemble des entrelaceurs périodiques [Cha04]. Ceux-ci sont utilisés en télécommunications pour améliorer la résistance au bruit, mais pas explicitement pour l'étalement de spectre. De plus la formulation en PCC conduit à une étude fréquentielle, qui est rarement effectuée pour les entrelaceurs.

L'application aux PCC de diverses structures de récepteur courantes en communications multi-utilisateurs [Ver98] a été étudiée dans [CR03] et [CR04] : récepteur à filtre adapté, récepteur décorrélateur et récepteur MMSE. Dans [Cri04], le problème de la turbo-synchronisation d'un système d'accès multiple utilisant les PCC est étudié. Les performances des PCC et du DS pour les communications multi-utilisateurs ont été comparées dans [RLT02] par rapport au nombre d'utilisateurs, avec $T = 2000$. Les estimations du TEB montrent des performances similaires pour un grand nombre d'utilisateurs et une légère supériorité des PCC pour un faible nombre d'utilisateurs, dans le cas d'un Rapport Signal sur Bruit (SNR) de 40 dB avec un bruit additif blanc gaussien. D'autres simulations (cf. *fig.* 2.1) effectuées avec un SNR plus proche de celui rencontré dans le tatouage (SNR=-10 dB) montrent une équivalence des performances des PCC et du DS-CDMA. Cependant, les paramètres utilisés lors de ces simulations (notamment le caractère gaussien du bruit introduit par le canal) reflètent un environnement de communication multi-utilisateur mais ne sont pas réalistes pour une application au tatouage.

2.1.2 Application au tatouage par étalement de spectre

Les Changements d'Horloge Périodiques et plus généralement les filtres LPTV n'ont jamais été utilisés jusqu'ici comme technique d'étalement au sein d'un algo-

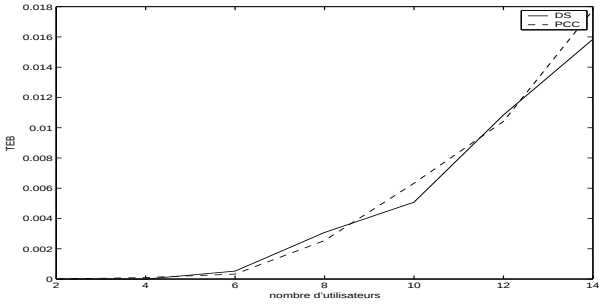

FIG. 2.1 – Comparaison entre PCC et DS en fonction du nombre d'utilisateurs : cas du canal AWGN, SNR=-10 dB

rithme de tatouage. On a cependant vu qu'à l'extrême ($T = N$) les PCC par permutations aléatoires correspondent à un entrelaceur aléatoire. Le cas particulier des permutations aléatoires a déjà été envisagé à divers niveaux. De nombreux auteurs les utilisent pour entrelacer le message dans une phase de pré-traitement afin d'améliorer la sécurité du tatouage [HPG99][KM03]. Leur intérêt est ici purement cryptographique, l'étalement étant assuré par une séquence pseudo-aléatoire. Cet entrelacement a également été appliqué à un motif reconnaissable visuellement (logo), dans le cadre d'une insertion non additive dans les moyennes fréquences de la DCT par blocs [HW99]. Furon et Duhamel proposent d'utiliser les permutations aléatoires pour l'entrelacement d'une information possédant un spectre coloré dans le cadre du tatouage asymétrique (cf. paragraphe 1.4.1) [FD03]. Il s'agit bien d'une utilisation pour l'étalement de spectre et la sécurité, mais l'information insérée est binaire (présence du tatouage ou non). Il ne s'agit pas de permuter par les bits du message eux-mêmes comme dans la suite de ce paragraphe. Enfin, la thèse de G.F. Elmasry est consacrée à l'étude de divers entrelacements non aléatoires 2D et 3D pour améliorer la détection des tatouages d'images et vidéos [Elm99]. Certains types d'entrelacements peuvent être vus comme des cas particuliers de PCC. Cependant, dans [Elm99] l'entrelacement est utilisé uniquement pour améliorer la robustesse aux attaques générant des rafales d'erreurs 2D, telles que le rognage de blocs ou la compression JPEG. L'entrelacement n'est pas aléatoire (les matrices sont même générées par récurrence) et est employé en plus de l'étalement DCT-DS à mise en forme NRZ.

L'utilisation des PCC dans le cadre du tatouage multiple impose de construire un ensemble de filtres :

1. étalant le spectre
2. inversibles
3. cryptographiquement sûrs
4. orthogonaux

La condition 1. impose une période T suffisamment grande. La condition 2. est difficile à réaliser pour des PCC quelconques. La condition 3. impose un caractère aléatoire à la fonction f. La condition 4. concerne le tatouage multiple. L'ensemble de ces contraintes et la nécessité d'une faible complexité des calculs nous a conduit à choisir les permutations aléatoires périodiques, qui satisfont les quatre contraintes.

Dans l'étape Sp de génération de $\mathbf{w}$ de l'algorithme proposé, on étale le spectre de $\mathbf{b}$ par un changement d'horloge périodique fondé sur une permutation aléatoire. La *fig.*2.3 montre l'effet d'une permutation aléatoire sur le spectre de l'image Lena. L'étape de réception Sp^{-1} consiste à appliquer la permutation inverse afin d'étaler le document support et les éventuels bruits additifs introduits lors de la transmission, et

de décoder le message.

PCC Mono-dimensionnels (1D-PCC)

1D-PCC utilise le format vectoriel (1.2). Afin d'atteindre un TEB raisonnable avec un SNR correspondant aux conditions du tatouage numérique, on introduit tout d'abord une redondance. Le principe d'attribuer à un bit d'information $l \in \{1, \ldots, L\}$ un ensemble $\mathcal{S}_l$ de $P = N/L$ pixels est classique (cf. paragraphe 1.2.2). Une mise en forme NRZ est impossible lorsque DWR est faible, *i.e.* P très grand : la période T de la permutation serait alors trop faible pour que le décodage d'un bit donné soit indépendant des sauts de moyenne locale. Inversement, avec une mise en forme aléatoire, on enlèverait toute corrélation au sein du document. Les performances des PCC seraient strictement identiques à celles de la technique DS, et le principe intuitif de périodicité perdrait son sens s'il était précédé d'une permutation aléatoire globale. Pour bénéficier de propriétés spatiales particulières du document précisées dans le paragraphe A.1.1, nous construisons donc $\mathcal{S}_l$ périodique modulo P : c'est la "mise en forme répétition". Le message résultant $\mathbf{b}^j$ est donc la concaténation de P répliques de $\mathbf{m}^j$:

$$b^j_{l+(p-1)L} = m^j_l, l \in \{1, \ldots, L\}, p \in \{1, \ldots, P\} \ . \tag{2.1}$$

$\mathbf{w}^j$ est obtenu en appliquant un PCC T_{1D}-périodique f_j (la clé secrète) à $\mathbf{b}^j$ (étape Sp). Le décodage S^{-1} consiste à appliquer la permutation inverse f_j^{-1} à $\mathbf{z}$, puis à moyenner les P échantillons correspondant à chaque bit du message initial. En négligeant les Interférences Multi-Utilisateurs (MAI), on a :

$$\begin{aligned} \hat{d}^j_l &= \frac{1}{P}\sum_{p=1}^{P}(f_j^{-1}(\mathbf{z}))(l + (p-1)L) \\ &= \psi m^j_l + \frac{1}{P}\sum_{p=1}^{P}(f_j^{-1}(\mathbf{x}+\mathbf{n}))(l + (p-1)L) \ . \end{aligned} \tag{2.2}$$

Supposons que les bits $+1$ et -1 soient équiprobables dans $\mathbf{m}^j$ et que L soit suffisamment grand. Le Théorème Central-Limite permet d'affirmer que pour une grande valeur de P, $\hat{d}^j_l$ est gaussien avec

$$E[\hat{d}^j_l] = \psi \mathbf{m}^j + \mu(\mathbf{y}) \text{ et } \mathrm{Var}(\hat{d}^j_l) = (\sigma^2_{\mathbf{y}} + \sigma^2_{\mathbf{n}})/\psi^2 P \ .$$

$\hat{\mathbf{d}}^j$ est alors une statistique suffisante pour le décodage. Soit $\hat{\mathbf{d}}^j = \hat{d}^j_l - \mu(\mathbf{y})$, où $\mu(\mathbf{y})$ est la moyenne de $\mathbf{y}$. La règle de décision correspondante est

$$\hat{\mathbf{m}}^j = [\mathrm{signe}(\hat{d}^j_l - \mu(\mathbf{y}))]_{l\in\{1,\ldots,L\}} = [\mathrm{signe}(d^j_l)]_{l\in\{1,\ldots,L\}} \ .$$

Ce récepteur correspond au filtre adapté [Ver98] à la "forme d'onde répétition" dont le support est de taille P (cf. 2.1.3).

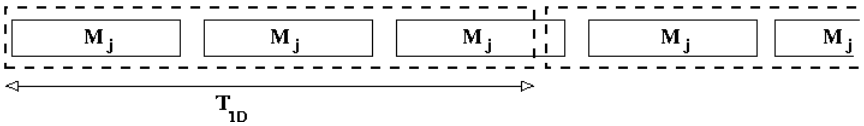

FIG. 2.2 – Application des PCC 1D sur le message à l'insertion

Étalement d'une image à la réception pour $T_{1\mathrm{D}} = 2^6, 2^{12}, 2^{18}$

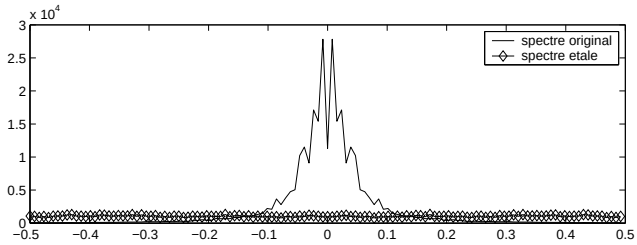

FIG. 2.3 – propriétés d'étalement des PCC (Lena)

PCC Bidimensionnels (2D-PCC)

1D-PCC peut être appliqué sur tout document numérique mis sous forme vectorielle. Dans le domaine du tatouage d'image, 2D-PCC utilise le format matriciel (1.1). Le tatouage à spectre étalé est le résultat de l'application successive à $\mathbf{b^j}$ d'une permutation f_j^1 sur les colonnes et d'une seconde permutation f_j^2 sur les lignes. Au décodage :

$$\hat{d}_l^j = \frac{1}{P}\sum_{p=1}^{P}((f_j^2)^{-1} \circ (f_j^1)^{-1}(\mathbf{z}))(\lceil l+(p-1)L\rceil)(\lfloor l+(p-1)L\rfloor) \qquad (2.3)$$

avec $l+(p-1)L = N_1\lceil l+(p-1)L\rceil + \lfloor l+(p-1)L\rfloor$. L'insertion et le décodage suivent le même principe que les 1D-PCC (cf. *fig.* 2.4). On s'attend à avoir des performances similaires à celles des 1D-PCC pour une période plus faible : l'association de deux PCC au décodage doit éliminer efficacement la corrélation spatiale entre les pixels (dans le cas d'une insertion dans la luminance) ou entre les coefficients de la DCT.

Étalement de l'image à la réception pour $T_{2\mathrm{D}} = 2^6, 2^{12}, 2^{18}$

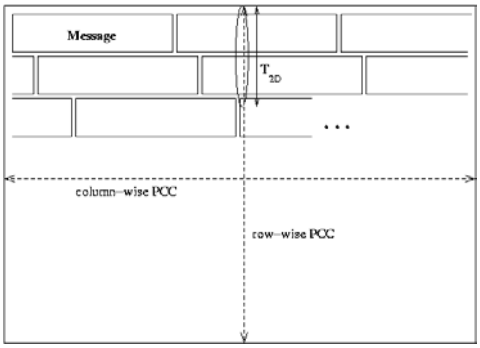

FIG. 2.4 – Application des PCC 2D sur le message à l'insertion

Insertion de logo

On a considéré jusqu'ici que **m** était un message binaire. On peut également insérer un logo, c'est-à-dire un message non binaire bidimensionnel. On a alors :

$$\mathbf{m} = [m_{k_1,k_2}]_{k_1 \in \{1,\ldots,K_1\}, k_2 \in \{1,\ldots,K_2\}} \quad .$$

Pour évaluer les performances, une mesure de similarité entre $\mathbf{m}$ et $\hat{\mathbf{m}}$ fréquemment utilisée est la corrélation normalisée [CMB02] :

$$\mathrm{NC} = \frac{\sum_{k_1=1}^{K_1} \sum_{k_2=1}^{K_2} m_{k_1,k_2} \hat{m}_{k_1,k_2}}{\sum_{k_1=1}^{K_1} \sum_{k_2=1}^{K_2} {m_{k_1,k_2}}^2} \quad .$$

L'insertion d'un logo est notamment utilisée en tatouage révélateur, où la déformation visuelle sur un logo d'un bloc donné permet d'identifier la localisation et le type d'une attaque. Dans le même but, le logo inséré peut également être une version très compressée de l'image elle-même ("auto-insertion" [FG99]).

En tatouage d'image, les PCC et en particulier 2D-PCC se prêtent bien à l'insertion de motifs 2D, non nécessairement binaires. La *fig.* 2.5 en donne un exemple. Les PCC effectuent une répartition aléatoire des pixels du logo sur toute l'image. La méthode se rapproche des travaux de Voyatzis et Pitas [VP98b] qui "mélangent" un logo par un automorphisme (bijection d'une grille de coordonnées sur une autre, de même taille). Son orbite (*i.e.* les espaces images lors d'applications successives de l'automorphisme) est périodique, ce qui permet de retrouver le logo initial après applications successives. Dans le même esprit, Tsekeridou *et al.* [TSN+01] proposent de mélanger les pixels par des systèmes chaotiques.

Signatures PCC

Il est possible d'adapter les PCC à la technique des "signatures" (cf. paragraphe 1.1.3). Supposons que $T = P$. Pour une séquence **b** fixée, nous proposons d'attribuer la permutation f_1 au bit 1, et f_{-1} au bit -1. Le tatouage est alors $w_k = b_{n-f_{m(\lceil n \rceil)}(n)}$. Il ne s'agit pas à proprement parler d'un PCC, mais plutôt du multiplexage de deux PCC. Le décodage utilise le maximum de vraisemblance entre les deux hypothèses. Cette solution est proche des signatures de type DS. Les signatures PCC se distinguent par la possibilité d'attribuer un sens au code **b** de départ (par exemple, une information sur le document) en plus de l'information transmise par la signature (par exemple, l'identifiant des usagers). On utilise ici le principe de détection de [CMB02] : un tatouage est

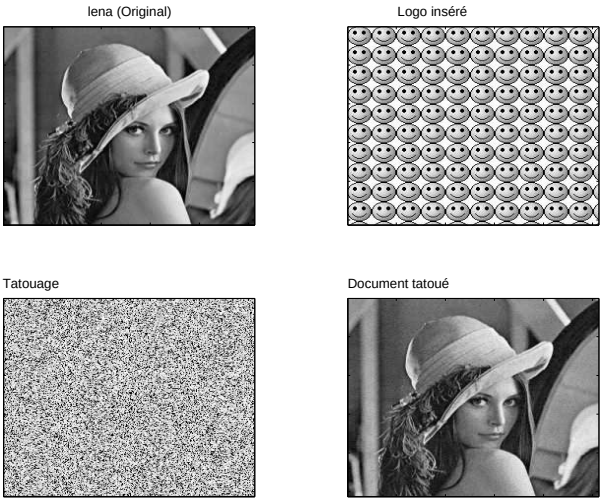

FIG. 2.5 – Insertion d'un logo avec les PCC

détecté si le message décodé a un sens.

Dans la suite de ce rapport, nous nous limiterons à l'insertion de messages binaires.

2.1.3 Performances théoriques face au bruit additif blanc gaussien

Problème de la détection

Le problème de la détection se formule comme un test d'hypothèses binaire :

$$H_1 : \mathbf{y} = \mathbf{x} + \mathbf{w}$$
$$H_0 : \mathbf{y} = \mathbf{x}$$

H_1 correspond à la présence d'un tatouage, H_0 à son absence. On utilise les mesures de performances suivantes :

$$P_d = \text{p(décider } H_1 | H_1) \quad \text{: probabilité de détection}$$
$$P_{\text{nd}} = \text{p(décider } H_0 | H_1) \quad \text{: probabilité de non-détection}$$
$$P_{\text{fa}} = \text{p(décider } H_1 | H_0) \quad \text{: probabilité de fausse alarme}$$

La détection est parfois reliée au décodage sous l'appellation "tatouage zéro-bit", dans lequel le TEB est la "probabilité totale d'erreur" TEB=$\frac{1}{2}(1 - P_d + P_{\text{fa}})$ [BDBT06].

Détecteur binaire : dans le cadre du tatouage par étalement de spectre et sous l'hypothèse AWGN, les décisions sur chaque bit sont indépendantes. On se ramène donc à une démodulation d_l^j, la dimension du problème étant réduite à un scalaire. Le détecteur de Neyman-Pearson est le plus souvent utilisé en tatouage numérique. En effet, la valeur maximale de P_{fa} a un sens pratique : elle se retrouve par exemple dans des cahiers des charges comme celui du SDMI. D'autres stratégies de détection, comme la minimisation du risque bayésien, seraient cependant possibles. Pour un bit

d'information et pour P suffisamment grand, le test de Neyman-Pearson correspond à :

$$H_1 : d_l^j \sim \mathcal{N}(\pm\psi, \frac{\sigma^2_{\mathbf{y}|H_1}}{P})$$
$$H_0 : d_l^j \sim \mathcal{N}(0, \frac{\sigma^2_{\mathbf{y}|H_0}}{P})$$

Le détecteur qui rend maximale la probabilité de détection P_d pour une valeur donnée de la probabilité de fausse alarme P_{fa} est :

$$\text{accepter } H_1 \text{ si} \quad \ln \frac{\text{p}(d_l^j|H_1)}{\text{p}(d_l^j|H_0)} > \eta$$

η étant un seuil calculable à partir de P_{fa} et de $\text{p}(d_l^j|H_0)$. Un résultat classique est [Bar05][PFCPG05] :

$$P_{\text{fa}} = Q\left(\frac{\sqrt{P}\eta}{\sigma_{\mathbf{y}|H_0}}\right) \text{ et } P_d = 1 - Q\left(\frac{\sqrt{P}(\psi-\eta)}{\sigma_{\mathbf{y}|H_1}}\right) \text{ où } Q(x) = \int_x^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du \ .$$

Ce résultat est confirmé en pratique par les courbes caractéristiques opérationnelles du récepteur (COR) donnant la probabilité de détection P_d en fonction de la probabilité de fausse alarme P_{fa}. La *fig.* 2.6 présente une application au tatouage d'image. Les performances de 2D-PCC face à une image naturelle y sont légèrement meilleures que celles de DS et 1D-PCC.

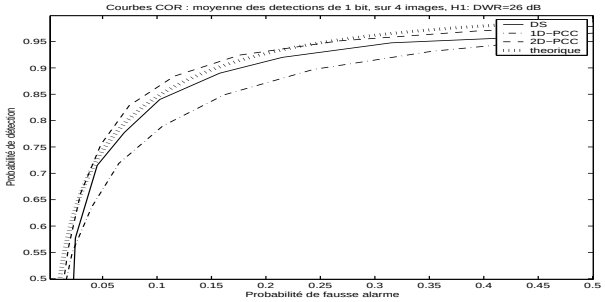

FIG. 2.6 – Courbe COR : détection de 1 bit, P=2621, DWR=26 dB

Détecteurs multi-symboles : lorsque le message comporte plusieurs bits (on parle alors de "détection-estimation conjointe" [Tre68]), plusieurs stratégies possibles sont discutées en annexe B.1.2. Nous y suggérons d'utiliser un "détecteur d'énergie" simple. Là encore, les performances des PCC face à un AWGN sont similaires à celles de DS.

Problème du décodage

Le récepteur à filtre adapté proposé dans le paragraphe 2.1.2 minimise la probabilité d'erreur sous les hypothèses de bruit gaussien et d'orthogonalité entre utilisateurs. Sous ces hypothèses, le récepteur optimal au sens des moindres carrés (MMSE) coïncide avec le filtre adapté. L'expression théorique du TEB [Ver98] est alors :

$$\text{TEB} = Q\left(\sqrt{\frac{P\sigma^2_{\mathbf{w}}}{\sigma^2_{\mathbf{x}} + \sigma^2_{\mathbf{n}}}}\right) = Q\left(\sqrt{\frac{P}{(\text{DWR} + 1/\text{WNR})}}\right) \ .$$

Le partage du débit dans le domaine spatial (**m** multi-bit) offre les mêmes performances que le partage en puissance (transmission multi-utilisateurs ou multi-porteuses) lorsqu'on néglige les interférences multi-utilisateurs (MAI). Les performances théoriques des PCC au décodage en prenant en compte les MAI sont présentées dans l'annexe B.1.3. Enfin, l'adaptation des PCC au tatouage multiplicatif est présentée dans l'annexe B.2.

2.1.4 Etude de la robustesse : application à l'image

Nous présentons ici un tableau récapitulatif des performances expérimentales des PCC pour le tatouage d'images naturelles. Les simulations complètes et commentées sont fournies dans l'annexe A.1. La propriété essentielle mise en avant par cette étude expérimentale est que les performances dans le domaine spatial sont très dépendantes de T (donc de la structure du PCC) et de L (donc de la structure de la mise en forme répétition), ce qui est particulièrement sensible pour l'algorithme 2D-PCC. Notons que DS est moins sensible à la mise en forme. En effet, pour DS une source de bruit est une zone support $\mathcal{S}_l$ de variance locale élevée. La moyenne locale n'a pas d'incidence car elle est annulée par le code. Pour les PCC, si la zone comporte un écart constant à la moyenne de l'image, le décodage sera faussé.

En effet, lors du décodage de 2D-PCC, Sp^{-1} estime m_l par la somme de P pixels pris parmi $\frac{N}{T^2}$ blocs. Supposons que $\mathbf{x}$ présente des blocs de taille $T_{bloc} \times T_{bloc}$ de pixels éloignés de la moyenne de l'image. C'est le cas pour la plupart des images naturelles considérées, des "objets" plus clairs ou plus foncés sont présents. La mise en forme répétition de **m** garantit que chaque support $\mathcal{S}_l$ est réparti sur ces différents blocs. En rajoutant un entrelacement aléatoire 2D-PCC, un choix de $T_{2D} < T_{bloc}$ conserve cette répartition. Des simulations effectuées sur un damier de blocs noirs et blancs de taille 64x64 ont montré que $T_{2D} = 64$ offre de très bons résultats au décodage, tandis que $T_{2D} = 128$ offre de mauvaises performances. A l'inverse, si $\mathbf{x}$ possède une forte activité locale, les performances s'améliorent quand T_{2D} augmente, ce qui est cohérent avec l'explication proposée. Nos efforts pour construire un modèle théorique plus précis des performances de 2D-PCC n'ont cependant pas abouti. La comparaison de performances sera donc expérimentale.

Le principe de dispersion du bruit sur tous les bits du message rejoint la problématique de l'entrelacement 2D [BBV98]. La construction d'entrelaceurs adaptés à l'image n'a cependant jamais été étudiée : le bruit considéré habituellement est composé de "rafales 2D" d'erreurs, c'est-à-dire sur des ensembles connexes de pixels. Les matrices d'entrelacement 2D sont construites de façon déterministe pour maximiser l'éloignement entre les éléments d'une rafale d'erreur 2D en sortie de l'entrelaceur. Les performances de 2D-PCC montrent que la combinaison d'une composante aléatoire dans l'entrelacement, de la mise en forme répétition et d'une prise en compte de la composition de l'image hôte est plus efficace qu'un entrelaceur déterministe. Les PCC avec mise en forme répétition effectuent un entrelacement aléatoire contrôlé qui permet de moyenner le bruit entre les supports des différents bits, et donc d'éviter les cas défavorables.

	L-DS	L-1D-PCC	L-2D-PCC	DCT-DS	DCT-1D-PCC	DCT-2D-PCC
dépendance à L	aucune	faible, ppcm(L,N) choisi grand	très dépendant, ppcm(L,N) choisi grand	aucune	faible	faible
dépendance à T	aucune	T choisi grand	très dépendant, T choisi faible	aucune	faible	faible
robustesse à **x**	similaires	similaires	similaires	similaires	similaires	similaires
tat. multiple	similaires	similaires	similaires	similaires	similaires	similaires
AWGN	similaires	similaires	similaires	similaires	similaires	similaires
débruitage	mauvaise	mauvaise (similaire à DS)	mauvaise (légèrement inférieur à DS)	moyen	légèrement meilleur que DCT-DS	légèrement meilleur que DCT-DS
compression	mauvaise	mauvaise (similaire à DS)	mauvaise (légèrement inférieur à DS)	moyen	légèrement meilleur que DCT-DS	légèrement meilleur que DCT-DS

2.2 Parcours de Peano-Hilbert de l'image

2.2.1 Présentation des courbes de remplissage d'espace

Le passage d'une image 2D x_{k_1,k_2} à un vecteur x_k se fait généralement en concaténant les lignes ou les colonnes. C'est cette technique que nous avons utilisée jusqu'à présent dans 1D-PCC. Cependant, d'autres parcours d'image sont possibles, notamment grâce aux courbes de remplissage d'espace (*space-filling curves*) introduites par Giuseppe Peano en 1890. Par exemple, la méthode "zig-zag" utilisée pour le parcours des coefficients de la DCT dans la compression JPEG est un parcours d'espace qui privilégie l'ordonnancement basses fréquences vers hautes fréquences. La technique de parcours d'espace la plus populaire est le parcours de Peano-Hilbert, qui procède récursivement [DCOM00]. Un exemple de parcours de Peano-Hilbert est présenté sur la *fig.* 2.7. Son intérêt est de conserver une plus grande corrélation entre pixels voisins qu'un parcours ligne par ligne, ce qu'on observe bien sur l'exemple de la *fig.* 2.8.

Curieusement, les courbes de remplissage d'espace n'ont que rarement été utilisées à notre connaissance dans le cadre du tatouage numérique, à l'exception d'une récente application à la stéganalyse dans [Wes05] et pour un étiquetage des pixels dans [CB01]. L'application qui se rapproche le plus des PCC est celle de [VP98a], où une séquence chaotique 1D est transformée en une séquence 2D par un parcours de Peano, dans le but d'imposer une structure passe-bas au tatouage. Le parcours de Peano n'a donc jamais été utilisé jusqu'à présent pour améliorer la robustesse des techniques d'étalement de spectre au bruit d'une image hôte. En effet, la plupart des auteurs utilisent DS avec une mise en forme aléatoire, qui décorrèle les échantillons correspondant à un même bit. Il n'y a donc pas d'intérêt à préserver une corrélation avant étalement. Les travaux de [Elm99] font exception : les auteurs y combinent une mise en forme NRZ avec un parcours d'image en "boustrophédon" (parcours ligne par ligne connexe) ou en "spirale", dans le but d'appliquer un entrelacement 1D. Cependant, les propriétés de corrélation de l'image ne sont pas utilisées car l'insertion est effectuée dans le domaine de la DCT. Leur démarche consiste uniquement à combattre l'attaque de rognage d'ensembles connexes de pixels de l'image.

A l'opposé, on a vu que les PCC avec mise en forme répétition pouvaient tirer profit

de la non-stationnarité de l'image et des corrélations entre blocs de pixels pour fournir une meilleure robustesse au bruit de l'image. Ils peuvent donc bénéficier des propriétés d'un parcours d'espace. De même que 2D-PCC offre de meilleures performances si les blocs de pixels similaires sont dispersés sur les supports de plusieurs bits, il est intéressant de regrouper les pixels corrélés dans un vecteur afin de mieux les disperser.

Plusieurs parcours classiques sont envisageables, pour le même type de performances : courbe de Peano ternaire, courbe de Moore ou encore courbe de Sipierski, qui partage un triangle isocèle en deux triangles homothétiques. Dans la suite, l'implantation utilisera la courbe de Peano-Hilbert, aussi appelée "courbe de Peano binaire" car elle partage récursivement un carré en quatre carrés. Une implantation possible du parcours de Peano-Hilbert est présentée ci-dessous.

---

si $n = 0$ peano$_n = [0, 0]$
sinon
$[u_0, v_0]$=peano$(n-1)$
$u = 0.5 * [-0.5 + v_0 - 0.5 + u_0 0.5 + u_0 0.5 - v_0]$
$v = 0.5 * [-0.5 + u_0 0.5 + v_0 0.5 + v_0 - 0.5 - u_0]$
peano(n)=[u,v]

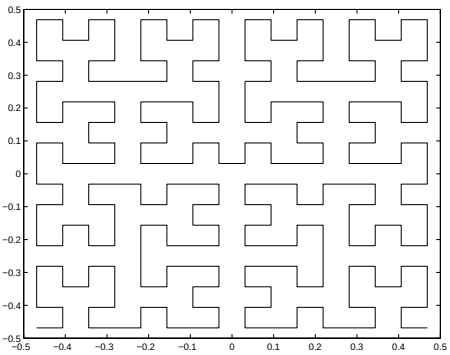

FIG. 2.7 – Exemple de parcours de Peano

2.2.2 Combinaison avec les PCC

Robustesse au bruit de l'image

Les performances expérimentales de la combinaison de 1D-PCC avec un parcours de Peano-Hilbert, appelée **PCC+Peano**, montrent une amélioration significative des performances par rapport à 1D-PCC et à DS. La contribution de l'image au bruit est divisée par un facteur allant jusqu'à 3, ce qui correspond à une multiplication de L par 3 sur la *fig.* 2.9. La *fig.* 2.11 montre que les bénéfices du parcours de Peano sont moindres après un préfiltrage de Wiener de l'image en réception (cf. partie 1.5.5).

Explication théorique des performances de PCC+Peano

Ces performances s'expliquent de la même façon que pour 2D-PCC (cf. *fig.* 2.12 à 2.14). Chaque bit l est associé à des échantillons $\mathcal{S}_l$ répartis régulièrement sur toute l'image. Une répartition aléatoire de ces échantillons sur l'image peut conduire à des

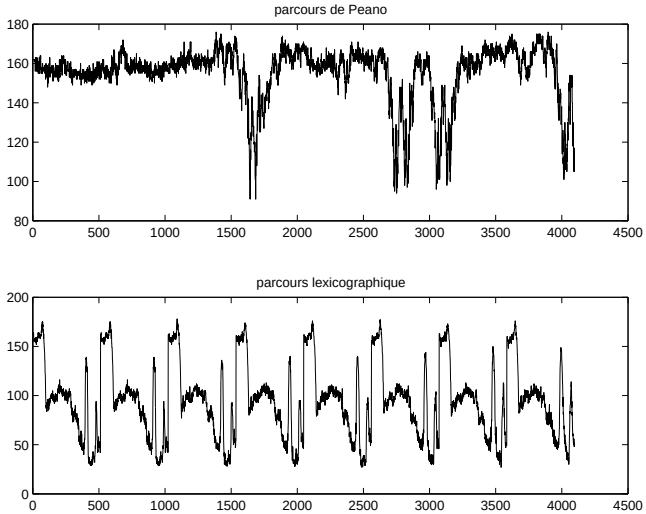

FIG. 2.8 – Intensité des 2^{12} premiers points de Lena selon le parcours

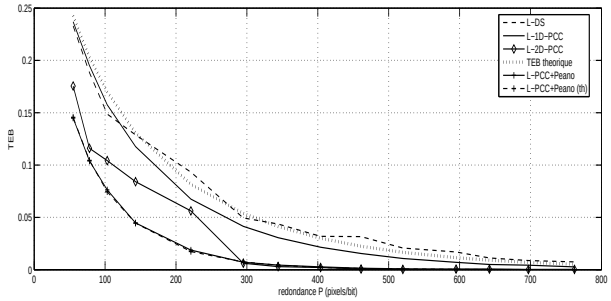

FIG. 2.9 – Performance au décodage par rapport à P, avec parcours de Peano

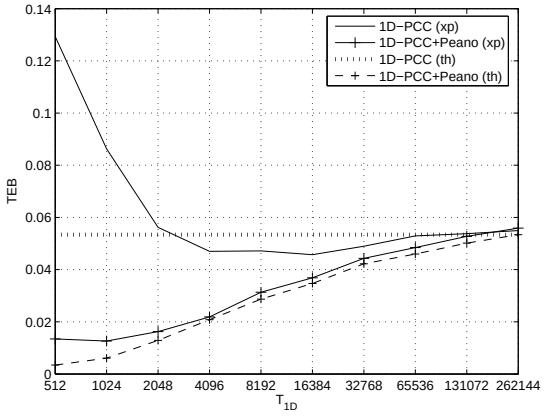

FIG. 2.10 – Performance au décodage par rapport à T, avec parcours de Peano

"paquets" d'échantillons situés sur un bloc de bruit fort, ce qui est un cas défavorable au décodage. A l'inverse, la mise en forme répétition assure une répartition régulière des échantillons entre les blocs, mais cette répartition n'est pas aléatoire. Si les permutations des PCC opèrent sur un bloc donné, on conserve un nombre limité d'échantillons du même bit dans le bloc permuté. Pour garantir une répartition moyenne du bruit et éviter les cas les plus défavorables, il faut donc permuter dans un bloc des points corrélés entre eux. Ceci a pour conséquence de minimiser l'écart à la moyenne au sein de

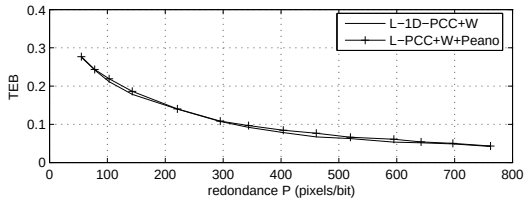

FIG. 2.11 – Performance au décodage par rapport à P, avec préfiltrage de Wiener, avec parcours de Peano, DWR=34dB

chaque bloc. Les écarts importants de variance du bruit entre chaque bloc permettent de répartir le bruit sur toute l'image. Si $\mathbf{x}$ était un bruit blanc, il n'y aurait pas d'influence de T. Cependant, $\mathbf{x}$ n'est que localement stationnaire.

L'analyse des blocs de l'image selon le parcours confirme l'influence du parcours de Peano et des PCC 2D. La *fig.* 2.15 montre que la variance moyenne de l'image sur chaque bloc de T_{1D} pixels consécutifs de l'image 1D diminue avec T_{1D} pour PCC+Peano et 2D-PCC, ce qui n'était pas le cas avec 1D-PCC. De plus, les écarts de moyenne entre chaque bloc sont importants. Ces propriétés sont atténuées par le pré-blanchiment de l'image avant décodage (cf. *fig.* 2.16).

Il est possible d'expliquer les performances de PCC+Peano en modélisant le bruit non plus comme une source gaussienne de variance $\sigma^2_{\mathbf{x}}$, mais comme N/T_{1D} sources gaussiennes parallèles de variance $\sigma^2_{\mathbf{x},k}$. Soit $\sigma^2_{\mathbf{x},k}$ la variance de $\mathbf{x}$ sur le bloc k. Alors dans le calcul du TEB théorique, on peut remplacer $\sigma^2_{\mathbf{x}}$ par $\mathcal{P}_{\mathbf{x},\text{PCC, Peano}} = \frac{T_{1D}}{N}\sum_{k=1}^{N/T_{1D}} \mathcal{P}_{\mathbf{x},k}$. Ce modèle n'est cependant valable que si T_{1D}/L est suffisamment grand. Si T_{1D} est trop petit par rapport à L, les performances expérimentales sont moins bonnes qu'en théorie (cf. *fig.* 2.9). En effet, on n'a plus un échantillon de $\mathcal{S}_l$ par bloc dans ce cas. Les simulations confirment ce raisonnement : il y a une adéquation parfaite de 1D-PCC+Peano avec sa courbe théorique sur la *fig.* 2.10.

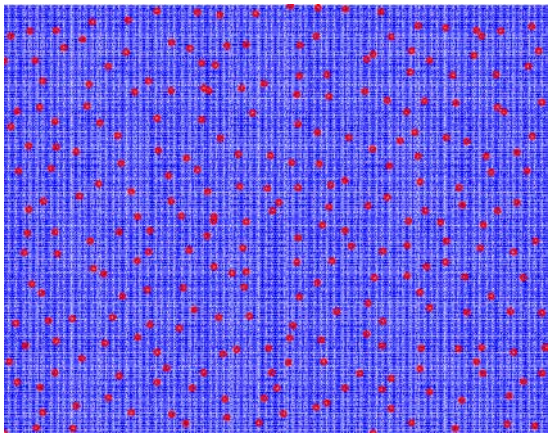

FIG. 2.12 – Répartition des échantillons associés à un bit : mise en forme répétition

2.2.3 Mise en forme PCC pour DS

On a envisagé jusqu'ici le cas des permutations aléatoires périodiques. Nous proposons désormais de pondérer ce PCC, par exemple par un code pseudo-aléatoire antipodal. Il s'agit alors d'un étalement DS qui bénéficie des avantages spatiaux de la mise en forme répétition. Nous appellerons cette technique "DS avec mise en forme

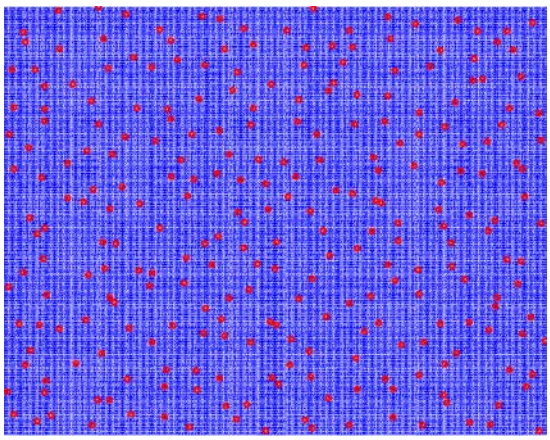

FIG. 2.13 – Répartition des échantillons associés à un bit : après PCC

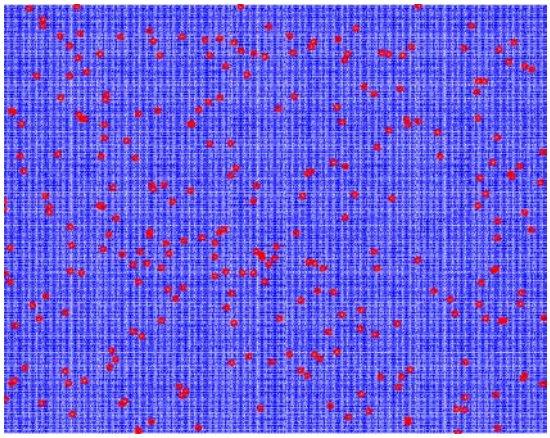

FIG. 2.14 – Répartition des échantillons associés à un bit : mise en forme aléatoire

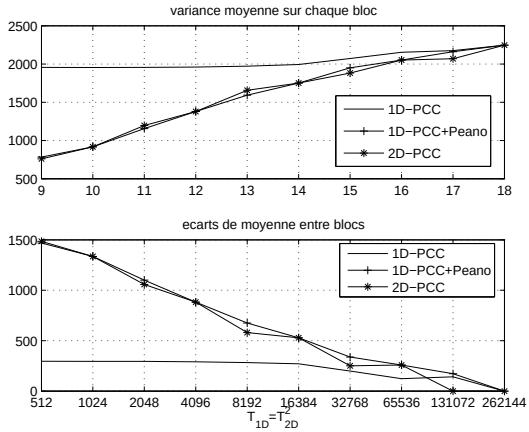

FIG. 2.15 – Dispersion du bruit au sein de chaque bloc, selon le parcours

PCC+Peano" (**PCC-DS+Peano**). Les simulations de la *fig.* 2.17 montrent cependant que les performances de cet algorithme sont similaires à celles de DS avec mise en forme répétition ou aléatoire : la modulation par un code pseudo-aléatoire élimine les particularités de la mise en forme PCC. Cela rejoint les explications du paragraphe 2.1.4 : DS est peu sensible à un fort écart à la moyenne sur $\mathcal{S}_l$. Ces résultats sont confir-

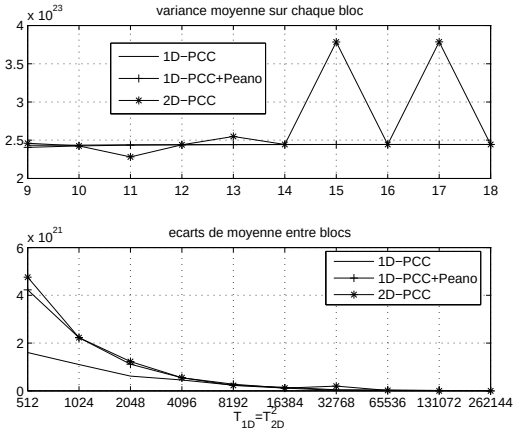

FIG. 2.16 – Dispersion du bruit au sein de chaque bloc, selon le parcours, après préfiltrage de Wiener

més par la *fig.* 2.18 sur la base de 44 images [Cit], où l'on note de plus que 2D-PCC fournit toujours les meilleurs résultats.

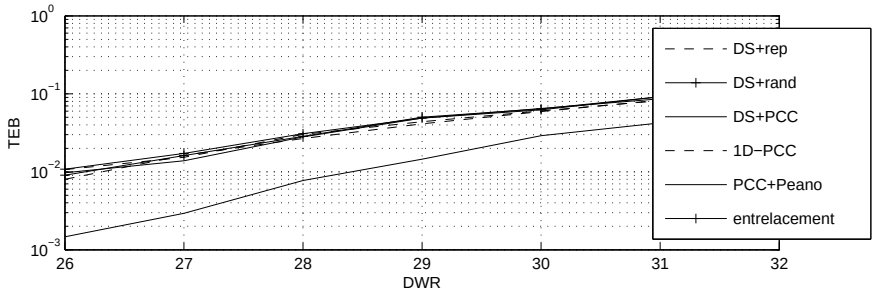

FIG. 2.17 – Robustesse au bruit de $\mathbf{x}$ en fonction de la mise en forme, $L = 100$

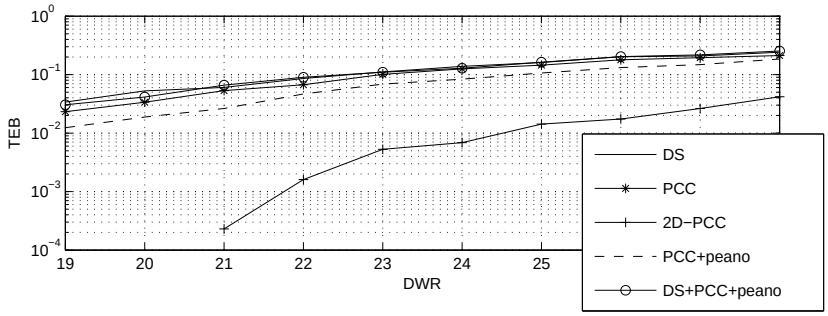

FIG. 2.18 – Robustesse au bruit de $\mathbf{x}$ en fonction de la mise en forme, 44 images, $L = 99$

2.3 Techniques de tatouage utilisant les filtres LPTV

Les changements d'horloges périodiques font partie d'un ensemble plus général de filtres, les filtres Linéaires Périodiques Variant dans le Temps (LPTV), qui présentent eux aussi sous certaines conditions des propriétés d'étalement de spectre, d'inversibilité et de sécurité. De plus, ils peuvent être construits de manière à prendre en compte le document hôte à l'insertion (tatouage informé), ou à améliorer l'imperceptibilité. Dans cette partie, on utilisera le parcours de Peano introduit dans la partie précédente.

2.3.1 Les filtres linéaires périodiques variant dans le temps (LPTV)

Définition et principe

Définition : un filtre Linéaire Périodique Variant dans le Temps (LPTV) est un filtre dont la réponse impulsionnelle est une fonction périodique du temps. Soit T cette période. Soient $\mathcal{F}^{\text{LPTV}}$ l'opération de filtrage, $h(n,k)$ la réponse impulsionnelle du LPTV, $\mathbf{v}$ la sortie du filtre et $\mathbf{u}$ son entrée ($\mathbf{v} = \mathcal{F}^{\text{LPTV}}(\mathbf{u})$). Alors

$$v_n = \sum_{k=-\infty}^{+\infty} h(n,k)u_{n-k}, \qquad h(n+T,k) = h(n,k) \ .$$

La fonction de transfert $H_n(\omega)$ du filtre définie par

$$H_n(\omega) = \sum_{k=-\infty}^{+\infty} h(n,k)e^{-ik\omega}$$

vérifie également : $H_n(\omega) = H_{n+T}(\omega)$.

Lien avec les PCC : les PCC sont un cas particulier de filtre LPTV. En effet, $v_n = u_{n-f(n)}$ donc $h(n,f(n)) = 1$ et $\forall k \in \mathbb{Z} \setminus f(n), h(n,k) = 0$. Leur fonction de transfert est $H_n(\omega) = e^{-if(n)\omega}$. Comme $f(n)$ est périodique de période T, H_n l'est aussi.

Décompositions : il existe plusieurs types de décomposition des filtres LPTV, qu'on peut retrouver dans [McL99], [AV00], [Vet89], [LB84] et [Cha04]. La décomposition en composantes modulatrices en sortie et la représentation état-espace ne seront pas utilisées. $U(\omega)$ et $V(\omega)$ désigneront ici les transformées de Fourier respectives de $\mathbf{u}$ et $\mathbf{v}$, et $U(z)$ et $V(z)$ leurs transformées en Z.

1. Décomposition en familles polyphases $\{\text{Tp}_k(z)\}_{k=1,\dots,T}$:

$$[u_{kT+i}] \rightarrow U_i(z) \triangleq \sum_{k=-\infty}^{\infty} u_{kT+i}z^{kT+i} \text{ et } V(z) = \sum_{k=0}^{T-1} \text{Tp}_k(z)U_k(z)$$

2. Décomposition en commutation de l'entrée et de la sortie (MIMO) :
on peut également décomposer la sortie $\mathbf{v}$ du filtre en familles polyphases

$$[v_{kT+i}] \rightarrow V_i(z) \triangleq \sum_{k=-\infty}^{\infty} v_{kT+i}z^{kT+i}$$

Alors le filtre peut s'écrire sous forme matricielle (cf. *fig.* 2.19) :

$$[V_i(z)] = [H_{i,j}(z)][U_j(z)] \tag{2.4}$$

3. Décomposition en composantes modulatrices en entrée $\{\mathrm{Tm}_k(z)\}_{k=1,\dots,T}$:

$$V(z) = \sum_{k=0}^{T-1} \mathrm{Tm}_k(z) U(W_T^k z)$$

ou encore (cf. *fig.* 2.20) :

$$V(\omega_v) = \sum_{k=0}^{T-1} \mathrm{Tm}_k(\omega_v) U(\omega_v - \frac{2\pi}{T}k)$$

Relation entre décompositions :

$$\left[U(W_T^k z)\right] = \frac{1}{T}\left[W_T^{ij}\right][U_k(z)]$$
$$[\mathrm{Tm}_k(z)] = \frac{1}{T}\left[W_T^{ij}\right][\mathrm{Tp}_k(z)]$$

Soit $\hat{X}$ le "vecteur modulateur" suivant :

$$\hat{U}(z) \triangleq [U(z)U(zW_N^1)\dots U(zW_N^{N-1})]^t$$

Alors on peut exprimer les filtres modulateurs sous forme matricielle (redondante) :

$$\hat{V}(z) = [\mathrm{Tm}_{j-i}(zW_N^i)]\hat{U}(z)$$

Soient F^+ et F^- les matrices des coefficients respectivement de la TFD et de la TFD inverse. La relation entre la décomposition MIMO et les filtres modulateurs est la suivante :

$$[H_{i,j}(z)] = \frac{1}{N} F^+ [\mathrm{Tm}_{j-i}(zW_N^i)] F^- \tag{2.5}$$

La décomposition en composantes modulatrices est particulièrement adaptée à l'analyse et la synthèse des filtres. Nous l'utiliserons donc afin de construire des filtres LPTV réalisant des objectifs précis sur le spectre de la sortie, à partir de la connaissance du spectre de l'entrée. Il est également possible d'implanter le filtrage dans le domaine fréquentiel dans certains cas simples. La décomposition en familles polyphases est quant à elle adaptée à une implantation efficace du filtre dans le domaine temporel.

Filtres LPTV et cyclostationnarité : on utilise classiquement en traitement du signal des processus stationnaires au second ordre. Un processus stationnaire X possède une moyenne constante : $\mu_X(n) = \mathrm{E}[x_k] \triangleq \mu_X$ (stationnarité au premier ordre) et une fonction d'autocorrélation K_X indépendante du temps m :

$$K_X(m,n) = E[x_m x_{m-n}^*] \triangleq K_X(n)$$

Or en sortie d'un filtre LPTV, $\mathrm{E}[v_k]$ dépend de n. Par exemple, supposons que $h(kT,0) = 1\ \forall k \in \mathbb{Z}$ et $h(n,k) = 0 \quad \forall (k,l) \in \mathbb{Z}^2 \setminus (T\mathbb{Z},0)$. Alors $\mathrm{E}[v_{kT}] = \mu_U$ mais $\mathrm{E}[v_k] = 0 \quad \forall n \in \mathbb{Z} \setminus T\mathbb{Z}$, donc $\mathbf{v}$ n'est pas stationnaire.

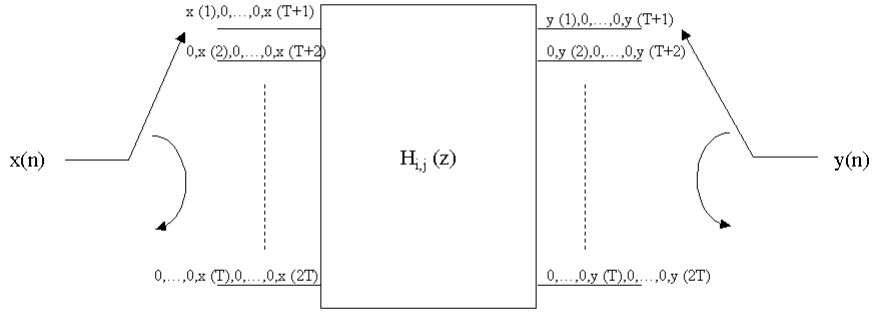

FIG. 2.19 – Décomposition MIMO en composantes polyphases d'un filtre LPTV

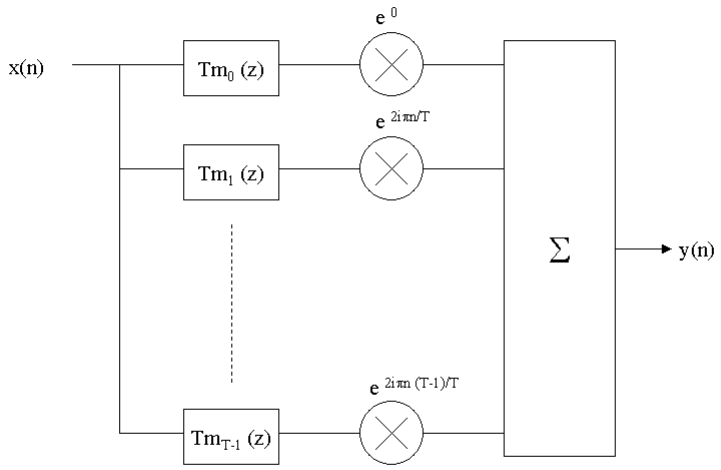

FIG. 2.20 – Décomposition filtres modulateurs en entrée d'un filtre LPTV

On appelle bispectre d'un processus non stationnaire X la fonction [AV00] :

$$S_X(\omega_v, \omega_u) \triangleq \frac{1}{2\pi} \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} K_X(m, m-n) e^{-i\omega_v m} e^{-i\omega_u n} \qquad (2.6)$$

De plus, on appelle représentation bifréquentielle d'un système linéaire variant dans le temps $h(n,k)$ la fonction [AV00] :

$$H(\omega_v,\omega_u) \triangleq \frac{1}{2\pi}\sum_{n=-\infty}^{+\infty}\sum_{k=-\infty}^{+\infty} h(n,n-k)e^{-i\omega_v n}e^{-i\omega_u k}$$

Dans le cas d'un filtre LPTV, on montre que

$$H(\omega_v,\omega_u) = F(\omega_v,\omega_u)\sum_{k=-\infty}^{+\infty}\delta(\omega_u-\omega_v+\frac{2\pi k}{T})$$

où

$$F(\omega_v,\omega_u) = \frac{1}{T}\sum_{k=0}^{T-1}\sum_{l=-\infty}^{+\infty} h(n,n-k)e^{-i\omega_v n}e^{-i\omega_u k}$$

La représentation bifréquentielle est donc distribuée le long de lignes parallèles d'équations $\omega_v-\omega_u = 2\pi k/T$. Ces lignes sont parallèles à la première diagonale $\omega_v=\omega_u$. On notera $\mathrm{Tm}_k(\omega) \triangleq F(\omega,\omega-2\pi k/T)$ la réponse impulsionnelle sur la ligne k. On a clairement $\mathrm{Tm}_{k+T}(\omega)=\mathrm{Tm}_k(\omega)$. Le système est donc complètement défini par les T premières diagonales (cf. *fig.* 2.21).

Un processus X est dit cyclostationnaire de cyclofréquence T si sa moyenne statistique et sa fonction d'autocorrélation sont périodiques de période T :

$$\mu_X(n)=\mu_X(n+T) \quad \text{et} \quad K_X(m,n)=K_X(m+T,n)$$

Là encore, le bispectre d'un signal cyclostationnaire est réparti sur des lignes parallèles à la première diagonale.

On peut montrer que la sortie $\mathbf{v}$ d'un filtre LPTV est cyclostationnaire, si l'entrée est stationnaire [Lac00]. On peut stationnariser $\mathbf{v}$ en considérant le processus $v_{n+\phi}$, où ϕ est une variable aléatoire uniformément distribuée sur $\{0,1,\ldots,T-1\}$ [LR02]. On montre également que le spectre de la version stationnarisée de $\mathbf{v}$ correspond à la première diagonale. Dans la suite, nous appellerons cette diagonale "spectre stationnaire" de $\mathbf{v}$. L'équation de filtrage peut alors s'écrire dans le domaine spectral [AV00] :

$$V(\omega_v) = \int_{-\pi}^{\pi} H(\omega_v,\omega_u)U(\omega_u)\mathrm{d}\omega_u \tag{2.7}$$

Puis pour un filtre LPTV, on se ramène dans le domaine fréquentiel à :

$$V(\omega_v) = \sum_{k=0}^{T-1} Tm_k(\omega_v)U(\omega_u-\frac{2\pi k}{T}) \tag{2.8}$$

On peut également écrire la densité de puissance de la version stationnarisée de $\mathbf{v}$ sous la forme [AV00][Cri04] :

$$S_v(\omega) = \sum_{k=0}^{T-1} |\mathrm{Tm}_k(\omega)|^2 S_u(\omega-\frac{2\pi k}{T}) \tag{2.9}$$

Le spectre de $\mathbf{u}$ est donc décalé dans le domaine fréquentiel sur T porteuses différentes et pondéré.

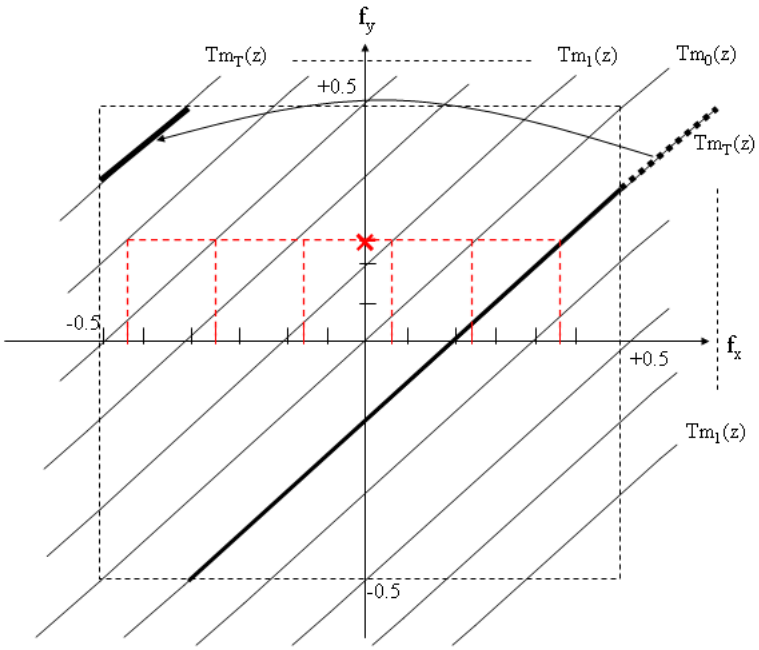

FIG. 2.21 – Diagramme bifréquentiel d'un filtre LPTV

Intérêt des filtres LPTV pour l'étalement de spectre

Les filtres LPTV ont une tendance naturelle à étaler le spectre, sans forcément le blanchir. En effet, pour un signal $\mathbf{u}$ à bande limitée dans $[-\frac{b}{2}, \frac{b}{2}]$, dès que T est suffisamment grande (par exemple $T = \text{Int}(2/b)$), il suffit que les filtres modulateurs $\text{Tm}_{j-i}(z)$ soient non nuls sur $f_u \in [-\frac{b}{2}, \frac{b}{2}]$ (correspondant à $\arg(z) \in \frac{2\pi}{T}[-1 + (j - i), 1 + (j - i)]$) pour que $V(f_v)$ soit non nul sur $f_v \in [-1, 1]$ car (cf. *fig.* 2.22) :

$$\hat{V}(z) = [\text{Tm}_{j-i}(zW_N^i)]\hat{U}(z)$$

Les filtres LPTV permettent ainsi de contrôler l'étalement sans blanchir. Il est également possible de construire le filtre afin de se prémunir des interférences à bande étroite, comme nous le proposons dans le paragraphe 2.3.3.

Un intérêt supplémentaire des filtres LPTV que nous entendons mettre en avant est que grâce à leur périodicité, ils bénéficient des propriétés déjà remarquées pour les PCC combinés au parcours de Peano lors du filtrage d'une image non-stationnaire.

Utilisation pour les télécommunications

Les filtres LPTV ont été appliqués à l'entrelacement, à l'égalisation aveugle et aux communications par étalement de spectre [McL99]. Cette dernière application a été l'objet des travaux de thèse de W. Chauvet [Cha04] au sein du laboratoire IRIT. Notamment, une famille de filtres LPTV inversibles dont la matrice des filtres modulateurs est circulante, appelée filtres LPTV convolutionnels, y est définie. Cependant, [Cha04] souligne la sensibilité au bruit de leur opération de filtrage inverse, ce qui rend les filtres LPTV convolutionnels peu adaptés au tatouage numérique. [Cha04] propose également

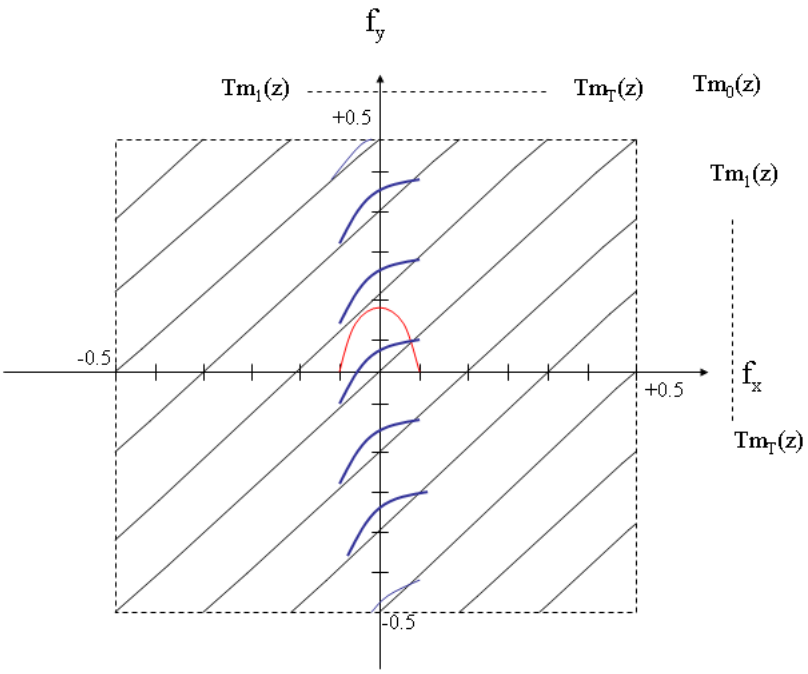

FIG. 2.22 – Capacités d'étalement d'un filtre LPTV sur sa représentation bifréquentielle

la construction d'une famille de filtres LPTV inversibles par analogie avec les bancs de filtres (LL-LPTV). Ces travaux ont été publiés dans [CCL+04] (système d'accès multiple fondé sur la condition d'orthogonalité et l'entrelaceur ligne/colonne). Ils ont été étendus à la synchronisation dans [CEL+04] et à l'égalisation dans [CRE05]. Les LL-LPTV ont été présentés dans [CLRD05]. Les travaux de thèse de B. Cristea [Cri04] utilisent principalement les LPTV comme un cadre théorique, et se concentrent plutôt sur les problèmes de turbo-synchronisation dans un système d'accès multiple utilisant les PCC.

Reformulation LPTV des techniques de tatouage existantes

Un certain nombre de techniques de télécommunication peuvent être reformulées comme des filtres LPTV, dans le but de les généraliser ou de les analyser avec les outils LPTV. C'est le cas de l'interpolation, de la décimation, de l'analyse multi-cadence [Cha04]. La même approche a moins de succès pour le tatouage. La reformulation est certes directe pour le CDMA :

$$v_k = u_k c_n = u(\lfloor n \rfloor) h(\lceil n \rceil)$$

où $h_k = \pm 1$. Mais dans le cas plus courant en tatouage (mise en forme aléatoire), DS n'introduit pas de périodicité. De plus, la reformulation est impossible pour LISS (cf. paragraphe 1.3.1), qui est un préfiltrage du tatouage. En effet, on doit prendre en compte un signal de bord et le filtrage est non linéaire. Quant au préfiltrage de Wiener (cf. paragraphe 1.5.5), il s'agit d'un filtre adaptatif non périodique. Notons enfin qu'il est possible de combiner les PCC fondés sur les permutations aléatoires avec la modulation par une courte séquence pseudo-aléatoire non secrète (mélange de PCC et de DS).

Cette méthode offre des résultats d'étalement et d'orthogonalité proches de ceux de DS. Cependant, il ne s'agit plus d'un PCC, car le code modifie les échantillons, mais bien d'un filtre LPTV de fonction de transfert $c_n e^{-if_n\omega}$. Le même résultat est obtenu pour la technique DS avec mise en forme PCC+Peano (cf. paragraphe 2.2.3).

Application au tatouage numérique

Les filtres LPTV n'avaient pas été appliqués au tatouage numérique jusqu'à présent. Plus généralement, l'insertion d'un tatouage cyclostationnaire n'a été évoquée que dans [dCTG02], à des fins de tatouage asymétrique, et en dehors du cadre des filtres LPTV. Un filtre LPTV appliqué au tatouage numérique doit être inversible. Deux approches existent. Dans la première, on se ramène à T filtres LIT inversibles, soit en prenant des constantes sur chaque intervalle $1/T$ de la matrice des filtres modulateurs, soit en imposant que le module de la réponse fréquentielle des modulateurs périodique aie comme période $1/T$ (filtres LPTV convolutionnels). Dans la seconde, on s'inspire des bancs de filtres à reconstruction parfaite. Pour éviter ces contraintes introduites par l'inversibilité, on pourrait également envisager des filtres LPTV quasi-inversibles (inversibilité non parfaite), puisque dans le cadre du tatouage ils seront utilisés en présence d'un très fort bruit.

Il est également important dans le cadre du tatouage d'images de définir la notion de spectre d'image. Sur une base d'images (tenant lieu de réalisations d'un processus aléatoire), il est possible de calculer le spectre 2D par transformée de Fourier 2D puis travailler sur les composantes de l'image dans une direction donnée. Une solution plus simple est, comme pour le cas des PCC, de transformer l'image en un signal mono-dimensionnel par un parcours d'image (cf. paragraphe 2.2). L'exemple des PCC nous montre que le parcours de Peano-Hilbert offre de meilleures performances que le parcours lexicographique, c'est donc la solution qui sera choisie dans la suite. Notons qu'il serait possible, comme cela a été fait pour les PCC, de construire des filtres LPTV bi-dimensionnels en travaillant d'abord sur les lignes, puis sur les colonnes. L'étude effectuée pour les PCC permet d'envisager de bonnes performances pour un tel système. L'étude ne sera pas approfondie en raison de son coût calculatoire.

Les filtres proposés seront appliqués à la transmission de messages avec une mise en forme répétition ou NRZ (cf. paragraphe 1.2.2). Le schéma de tatouage sera toujours additif :

$$\mathbf{y} = \mathbf{x} + \mathcal{F}^{\mathrm{LPTV}}(\mathbf{b}) \tag{2.10}$$

et à la réception :

$$\hat{\mathbf{b}} = \mathcal{F}^{\mathrm{LPTV}^{-1}}(\mathbf{z}) \tag{2.11}$$

2.3.2 Application de filtres LPTV existants au tatouage

Dans cette partie, on présente deux systèmes de filtres LPTV inversibles : les filtres LPTV sans perte (LL-LPTV), déjà utilisés dans le domaine des télécommunications, et les filtres LPTV inversibles à composantes modulatrices constantes (mod-LPTV). Leurs capacités d'étalement du spectre d'une image sont décrites. Enfin, on propose une variante de mod-LPTV possédant des propriétés d'orthogonalité.

Filtres LPTV sans perte (LL-LPTV) : analogie avec les bancs de filtres

Les filtres LPTV n'ont jamais été utilisés dans le cadre du tatouage. Par contre, les bancs de filtres sont fréquemment utilisés, en particulier dans tous les algorithmes

utilisant la transformée en ondelettes. Les bancs de filtres y sont utilisés comme pré- et post-traitements à l'insertion (tatouage dans un domaine transformé). Le principe d'insérer le tatouage en modifiant la fonction de transfert de certains filtres de ces bancs est moins fréquemment utilisé. On peut notamment citer [Z. 05a][Z. 05b] qui propose d'utiliser des "bancs de filtres à attribution de zéros" dans le cadre d'un algorithme de tatouage dans le domaine des ondelettes. Ces bancs de filtres sont les Filtres Miroirs en Quadrature à reconstruction parfaite. Le secret réside dans l'emplacement des zéros de la fonction de transfert de filtres passe-haut. La technique proposée dans la suite diffère de ces approches.

Définition des LL-LPTV : les filtres LPTV sans perte ou *Loss-Less* (LL-LPTV) ont été proposés par W. Chauvet ([Cha04], p.49). Une équivalence existe entre filtres LPTV et bancs de filtres [Cha04]. Cette analogie sert à construire les LL-LPTV à partir d'une famille de bancs de filtres à reconstruction parfaite introduite dans [Vai87]. Ils permettent de reconstruire un signal d'entrée sans distorsion d'amplitude ni de phase, en introduisant un simple retard. Cette famille de filtres LPTV a été appliquée aux télécommunications dans [Cha04]. Dans le cas d'un canal AWGN, les résultats expérimentaux de [Cha04] sont proches de ceux du CDMA. Le cas multi-utilisateurs n'est pas évoqué, et les performances théoriques n'ont pas été établies. Dans le cas plus particulier de la réduction de rapport entre la puissance crête et la puissance moyenne du signal (PAPR) dans les systèmes de communication à division fréquentielle orthogonale (OFDM), les LL-LPTV peuvent offrir de meilleures performances que les techniques classiques ([Cha04], p.119).

Les LL-LPTV sont construits à partir de matrices "sans perte" (stables et unitaires) construites de sorte que les puissances en entrée, en sortie du filtre et en sortie du filtre inverse, soient égales ($P_{\mathbf{u}} = P_{\mathbf{v}} = P_{\hat{\mathbf{u}}}$). Soit A une matrice de taille TxT construite à partir de vecteurs aléatoires et formant une base orthonormale (éventuellement après orthonormalisation de Gram-Schmidt). A est alors unitaire, donc inversible et son inverse est la transconjuguée de A. $H(z)$ désignant la matrice polyphase, on introduit p retards grâce à des vecteurs aléatoires C_i :

$$H(z) = H_p(z)H_{p-1}(z)\dots H_1(z)A \text{ avec } H_i(z) = (I - z^{-1})C_i$$

Sans les retards ($p = 0$), l'implantation est simple : (2.4) devient $[V_i(z)] = A[U_j(z)]$. Par linéarité de la transformée en Z, l'implantation du filtre se réduit à un calcul matriciel MIMO appliqué aux composantes polyphases :

$$[v_{kT+i}] = A\,[u_{kT+i}] \tag{2.12}$$

Si $p = 0$, le filtre LPTV inverse est $[u_{kT+i}] = A^{-1}\,[v_{kT+i}]$. On se limitera dans l'implantation à une matrice A composée de réels de moyenne nulle. Dans [Cha04], l'utilité des retards n'est pas étudiée. Dans le cadre du tatouage et avec un choix aléatoire de A et des C_i, nos simulations ont montré que les retards ne bénéficient pas à la robustesse.

Application au tatouage d'images numériques : les capacités d'étalement des LL-LPTV dans le cadre du tatouage d'image sont montrées sur les *fig.* 2.23 à 2.26. La *fig.* 2.24 montre l'effet d'un LL-LPTV sur le spectre d'une image. Le spectre en sortie est clairement étalé, mais n'est pas plat si T est faible. Plus la redondance P est grande, moins le spectre en sortie est plat. La *fig.* 2.25 montre l'étalement par LL-LPTV du spectre d'un message à mise en forme NRZ. A la réception, les interférences de l'image font apparaître des pics en dehors de la bande spectrale du message.

La mise en forme a une grande influence. Notamment, avec la mise en forme répétition, le spectre de **b** est presque plat (visible sur le périodogramme cumulé). Sur le périodogramme entier, on voit apparaître les raies correspondant aux répétitions. Sur la *fig.* 2.26, on observe que si le ppcm(L,T) est grand, LL-LPTV étale bien le spectre du message original. Par contre, si L est multiple de T, les raies spectrales dues à la répétition (périodogramme non cumulé) se retrouvent après étalement. Même si en dehors de ces raies, le message est étalé, ceci est une situation à éviter. La mise en forme NRZ est quant à elle très sensible aux variations locales de l'image, notamment de sa moyenne sur un bloc de taille T. Elle devrait donc être précédée d'un entrelacement NxN du signal, ce qui aurait cependant pour effet d'éliminer toutes les corrélations locales qui nous intéressent ici. La mise en forme NRZ doit donc être accompagnée d'une stationnarisation du signal hôte (préfiltrage de Wiener au décodage, cf. paragraphe 2.4.1).

Performances théoriques : d'après l'équation (2.12), v_k est une combinaison linéaire de variables gaussiennes indépendantes. Par orthonormalité de A, sa variance est $\sigma^2_{\mathbf{v}} = \sum_{k=0}^{T-1} |a(k,l)|^2 \sigma^2_{\mathbf{u}} = \sigma^2_{\mathbf{u}}$. Les calculs du paragraphe 2.1.3 sont valables en remplaçant la fonction PCC f par le filtre LPTV. Les performances théoriques en présence de bruit AWGN sont donc identiques à celles de DS et PCC. Dans le cas où le bruit est une image, l'évaluation du TEB ne peut se faire qu'expérimentalement.

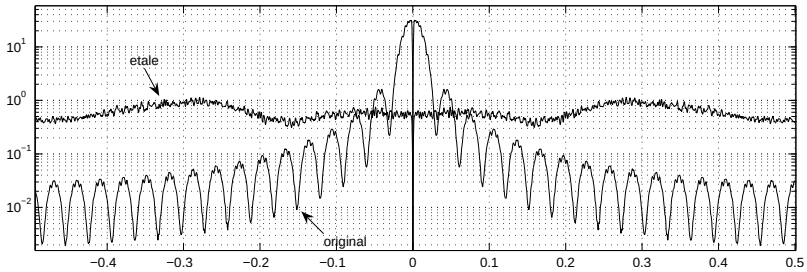

FIG. 2.23 – Étalement du spectre d'un signal NRZ par le LL-LPTV, $T = 8$, $p = 3$

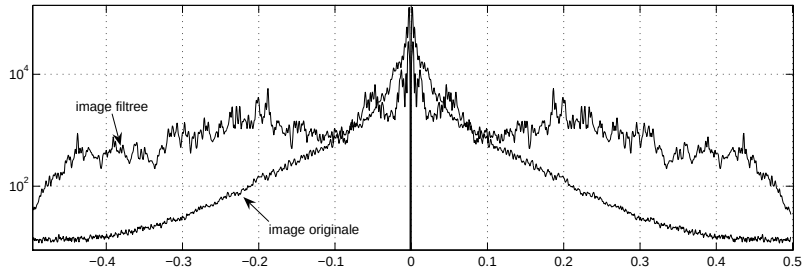

FIG. 2.24 – Étalement du spectre de l'image Bateaux par le LL-LPTV inverse, $T = 128$, $p = 3$

Filtres LPTV inversibles à composantes modulatrices constantes (mod-LPTV)

Définition de mod-LPTV : dans ce paragraphe, nous proposons de construire des filtres LPTV inversibles inspirés d'un article de Rohlev et Loeffler [RL87] qui n'a pas à notre connaissance été exploité dans le cadre des télécommunications. Les auteurs proposent, afin de construire un filtre LPTV inversible, d'utiliser des composantes modulatrices constantes par morceaux (cf. *fig.* 2.27).

Soit A une matrice de taille TxT. On pose $\mathrm{Tm}_{k-l}(e^{i2\pi f}) = a(k,l)$, avec $f \in$

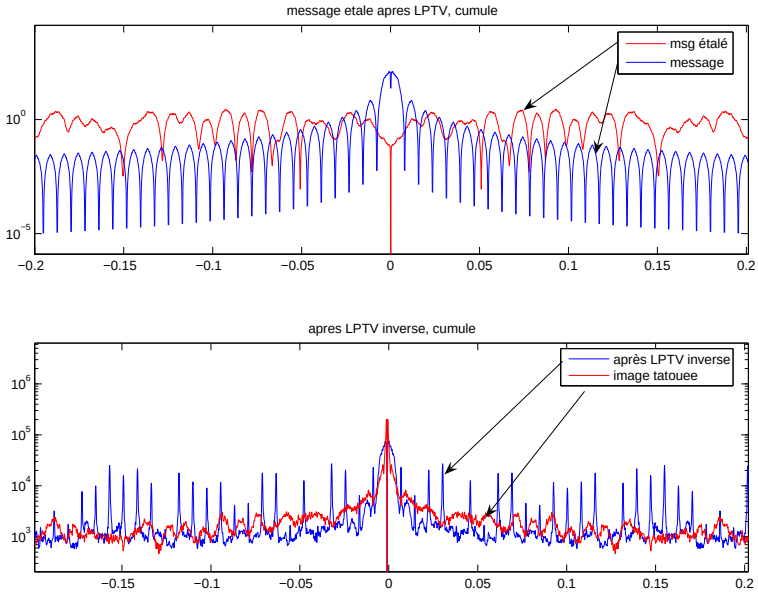

FIG. 2.25 – Influence spectrale de LL-NRZ-LPTV et de son inverse, DWR=5 dB, $L = 2048$, $T = 128$

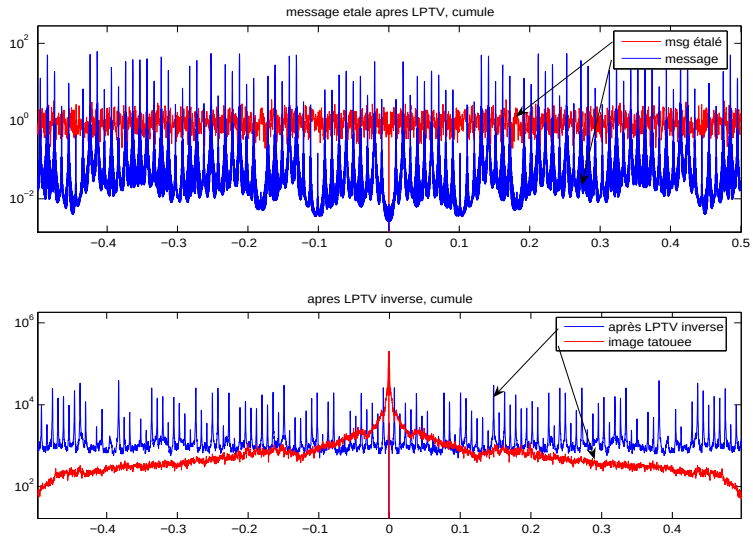

FIG. 2.26 – Influence spectrale de LL-LPTV et de son inverse, DWR=28 dB, $L = 100$, $T = 128$

$[\frac{k}{T} - \frac{1}{2}, \frac{k+1}{T} - \frac{1}{2}]$. L'équation de filtrage (2.8) s'écrit donc :

$$V(\omega_v) = \sum_{l=0}^{T-1} a(k, k-l) U(\omega_v - \frac{2\pi}{T} l) \tag{2.13}$$

ou encore, en notant V_k et U_k les régions de $V(\omega_v)$ et $U(\omega_v)$ comprises dans $2\pi[\frac{k}{T} -$

$\frac{1}{2}, \frac{k+1}{T} - \frac{1}{2}]$:

$$V_k = \sum_{l=0}^{T-1} a(k,l)U_l \tag{2.14}$$

On a donc une relation matricielle : $V = AU$. Il suffit que A soit inversible pour que le filtre soit inversible avec $U = A^{-1}V$. Dans [RL87], il est proposé d'approcher la fonction de transfert à synthétiser en construisant un filtre RIF pour chacune des fonctions de transfert des T filtres modulateurs, puis d'obtenir les coefficients de A par résolution d'un système linéaire. Pour que $\mathbf{v}$ soit réel, il faut que V soit paire. Comme U est paire, il faut donc que A soit symétrique par rapport à son centre : $a(T/2-k, T/2-l) = a(T/2+k, T/2+l)$.

Algorithme de tatouage proposé : l'algorithme de tatouage proposé consiste à générer aléatoirement une matrice A composée de réels, de taille TxT, puis à l'orthonormaliser. Le filtrage et le filtrage inverse des relations (2.10) et (2.11) sont ensuite réalisés dans le domaine fréquentiel, d'après les relations ci-dessus. Expérimentalement, nous avons constaté que l'orthogonalité de A (non requise par le principe initial) est très importante pour réduire la sensibilité du LPTV inverse au bruit de l'image non stationnaire. Comme il est difficile de construire une matrice orthonormale symétrique par rapport à son centre, il est préférable en pratique de relâcher la contrainte de symétrie de A. On ignore donc les composantes complexes des signaux en sortie du filtre.

Il est important de noter que ceci ne nuit pas à la détection. La perte de la moitié de la puissance du signal est compensée à l'insertion par le réglage de la force d'insertion. Au décodage, le bruit est autant affecté que le tatouage. De plus, le filtre est toujours inversible car dans notre cas, $\mathrm{Re}(\mathcal{F}^{-1}(\mathrm{Re}(\mathcal{F}(\mathbf{u})))) \propto \mathbf{u}$. En effet, on peut décomposer toute matrice A en une matrice S_1 symétrique par rapport à son centre et une matrice S_2 antisymétrique par rapport à son centre[1]. Le filtrage LPTV avec S_1 conduit à un signal réel, celui avec S_2 conduit à un signal imaginaire pur. Comme de plus A est orthonormale, $A^{-1} = S_1^t + S_2^t$. Le fait de ne conserver que la partie réelle du signal en sortie du filtre LPTV et de son inverse revient donc à filtrer avec S_1 puis S_1^t. Enfin, on peut montrer que $S_1 S_1^t = \mathrm{Id}/2$ si A est issue de l'orthonormalisation de Gram-Schmidt d'une matrice symétrique par rapport à son centre. A un facteur de puissance près, le filtre construit avec S_1 est donc inversible.

Réponse impulsionnelle : avec la matrice A utilisée, chacune des T réponses impulsionnelles correspondant aux fonctions de transfert $\mathrm{Tm}_k(z)$ a l'enveloppe d'un sinus cardinal de largeur de lobe T [RL87]. On peut recalculer ce résultat à l'aide de la formule (2.5) entre filtres modulateurs et polyphases. Il est confirmé expérimentalement par la *fig.* 2.28. L'énergie du signal n'est donc pas déplacée de façon importante temporellement. Pour LL-LPTV on voit clairement sur la *fig.* 2.29 que les filtres modulateurs sont RIF. Cette propriété est importante pour profiter des propriétés des filtres LPTV pour la dispersion du bruit de l'image. Malgré l'ajout de "trous", on retrouvera ce résultat expérimentalement pour ZI-LPTV (cf. partie 2.3.3). La méthode de construction de mod-LPTV permet de modifier l'allure temporelle du signal en introduisant des retards de type $k_a e^{i\omega a}$ dans A à la place d'une constante [RL87].

Performances théoriques : l'équation (2.9) permet de calculer analytiquement le

[1] prendre $S_1(T/2+l, T/2+k) = (A(T/2+l, T/2+k) + A(T/2-l, T/2-k))/2$ et $S_2(T/2+l, T/2+k) = (A(T/2+l, T/2+k) - A(T/2-l, T/2-k))/2$

spectre de $\mathbf{v}$ à partir de celui de $\mathbf{u}$. Dans le cas d'un bruit blanc gaussien, $S_u = \frac{1}{2\pi}$ et

$$S_v(\omega) = \frac{1}{2\pi} \sum_{k=0}^{T-1} |\mathrm{Tm}_k(\omega)|^2 \tag{2.15}$$

Dans le cas de mod-LPTV, il existe $l \in \{0, \ldots, T-1\}$ tel que $\omega \in [\frac{l}{T}, \frac{l+1}{T}]$ et

$$S_v(\omega) = \frac{1}{2\pi} \sum_{k=0}^{T-1} |a(k,l)|^2$$

Si A est orthonormale, on a donc $S_v(\omega) = \frac{1}{2\pi}$ et $\mathbf{v}$ est un bruit blanc. Cette propriété est aussi valable pour le filtre inverse. Les performances théoriques en présence de bruit AWGN sont donc identiques au cas DS, PCC et LL-LPTV. Notons que l'étude sur un bruit blanc élude les particularités de l'implantation réelle du filtre. Lorsque le bruit est une image hôte, l'influence théorique du bruit est difficile à calculer et est évaluée expérimentalement.

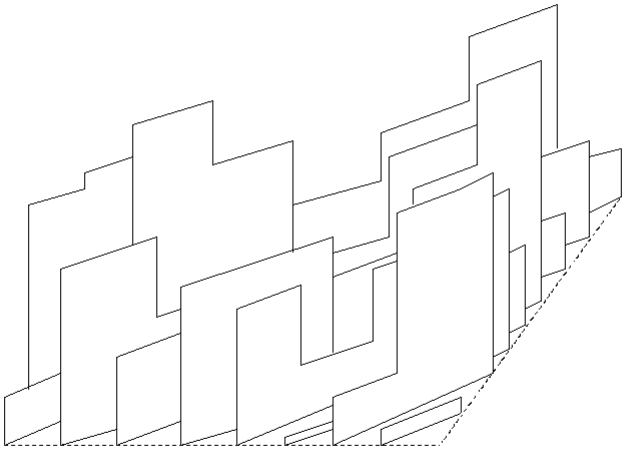

FIG. 2.27 – Diagramme bifréquentiel d'un filtre LPTV à filtres modulateurs constants par morceaux

2.3.3 Filtres LPTV et annulation des interférences de l'image

La formulation modulateur d'un filtre LPTV permet d'imposer des contraintes sur le spectre du signal en sortie, si le spectre de l'entrée est connu. Contrairement à la plupart des applications qui imposent des contraintes strictes au signal de sortie, le domaine du tatouage offre une grande liberté dans la conception du filtre. Dans cette partie, deux techniques sont proposées. La première a pour objectif de construire un

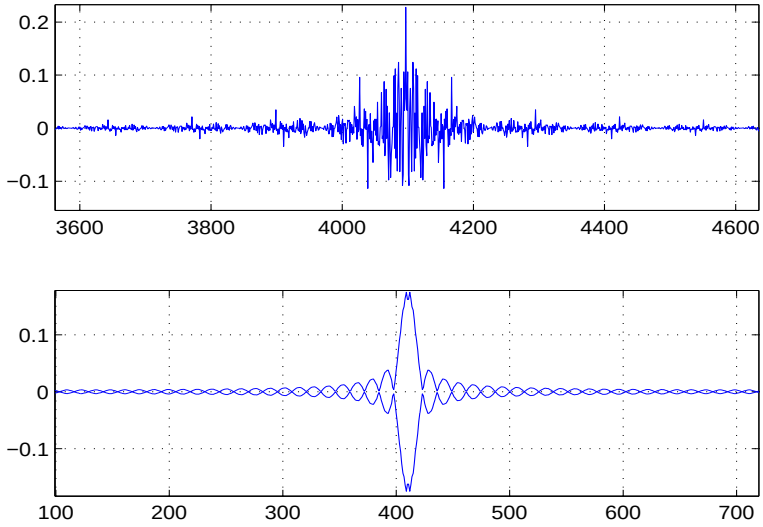

FIG. 2.28 – mod-LPTV : (haut) : exemple de réponse impulsionnelle d'un filtre modulateur, (bas) : enveloppe des filtres

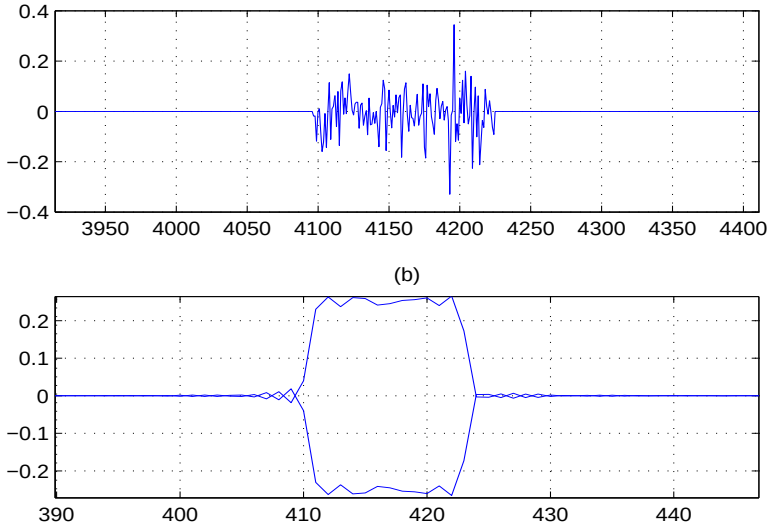

FIG. 2.29 – LL-LPTV : (a) : exemple de réponse impulsionnelle d'un filtre modulateur, (b) : enveloppe des filtres

décodeur peu sensible aux interférences de l'hôte, afin d'améliorer le TEB. La seconde a pour objectif d'imposer une contrainte perceptuelle sur le tatouage généré par le filtre LPTV. Les deux méthodes sont des extensions de mod-LPTV.

Méthode proposée (ZI-LPTV)

Nous proposons de construire un décodeur peu sensible au bruit de l'image. Cela revient à imposer une propriété au filtre LPTV *inverse*. L'idée est qu'en sortie du filtre

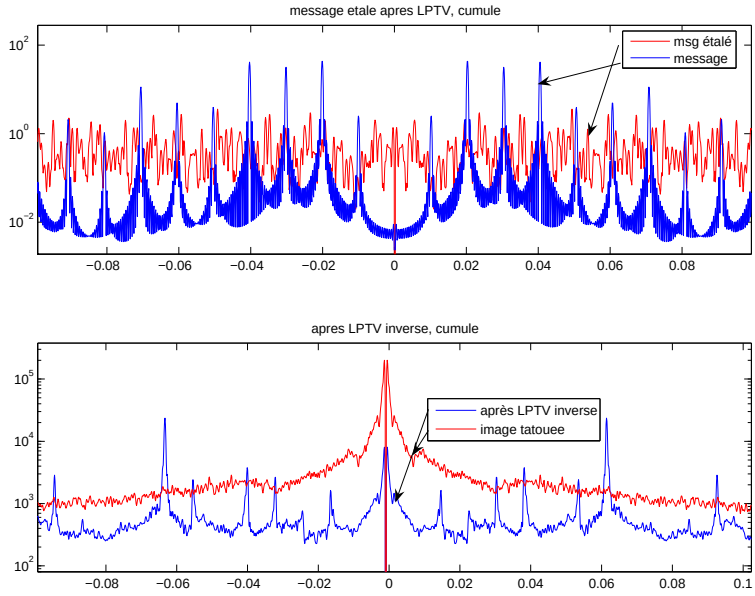

FIG. 2.30 – Influence spectrale de mod-LPTV et de son inverse, DWR=28 dB, $L = 100$, $T = 128$

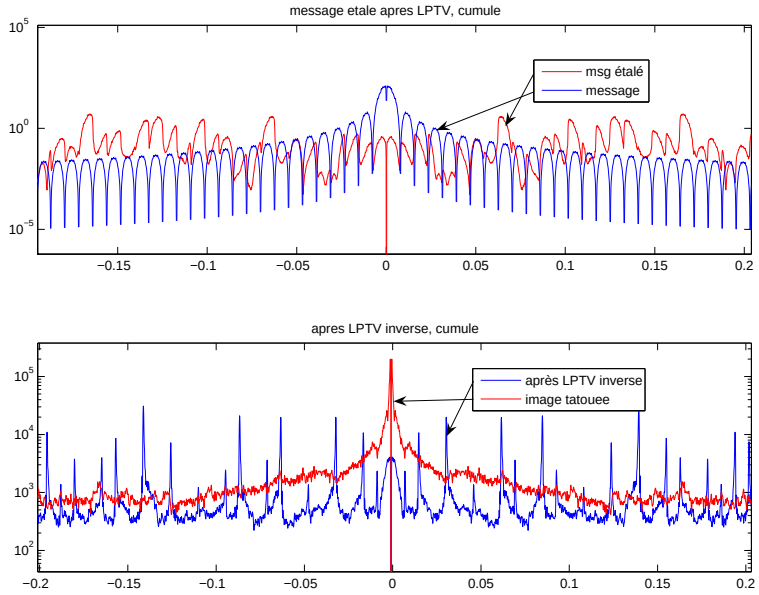

FIG. 2.31 – Influence spectrale de mod-NRZ-LPTV et de son inverse, DWR=15 dB, $L = 2048$, $T = 128$

LPTV inverse, la contribution du filtrage du document hôte soit nulle dans les composantes fréquentielles où se situe le message. Nous choisirons une mise en forme NRZ, afin que le message décodé soit à bande limitée comprise dans $[-\frac{b}{2}, \frac{b}{2}]$. En l'absence d'autre contrainte que l'inversibilité du filtre, nous proposons d'annuler complètement

la contribution de $\mathbf{x}$ dans cette bande. La contrainte est donc :

$$V(\omega_v) = 0 \quad \text{si} \quad f_v \in [-\frac{b}{2}, \frac{b}{2}]$$

La *fig.* 2.32 montre les contraintes à imposer sur les filtres modulateurs du LPTV inverse pour imposer à V de s'annuler dans les basses fréquences si $\mathbf{u}$ (le document hôte) a une bande limitée. Afin d'imposer cette contrainte, on étend l'algorithme mod-NRZ : les filtres modulateurs sont constants par morceaux et le filtrage est matriciel. Dans un premier temps, on génère aléatoirement une matrice orthonormale A. Puis la contrainte correspond à imposer

$$\begin{aligned} a^{-1}(k,l) = 0 \quad \text{pour} \quad & k \quad \text{situé dans la bande de fréquence de l'hôte à éliminer} \\ \text{et} \quad & l \quad \text{situé dans la bande visée contenant le spectre de } M \end{aligned} \tag{2.16}$$

Le spectre d'un signal NRZ antipodal de période P est $P\sin^2(\pi f P)\mathrm{sinC}^2(\pi f P)$. 99% de sa puissance est contenue dans le lobe principal. Les colonnes concernées sont donc

$$l \in \{\frac{T}{2} - (\text{Int}(\frac{T}{P}) + 1), \ldots, \frac{T}{2} + (\text{Int}(\frac{T}{P}) + 1)\}$$

En sortie, on applique un filtre passe-bas sur $[-\frac{b}{2}, \frac{b}{2}]$. On nommera cette technique *Zero-Insertion* LPTV (ZI-LPTV). Cette appellation ne doit pas être confondue avec les filtres à attribution de zéro de [Z. 05a], où les zéros correspondent au secret du tatouage.

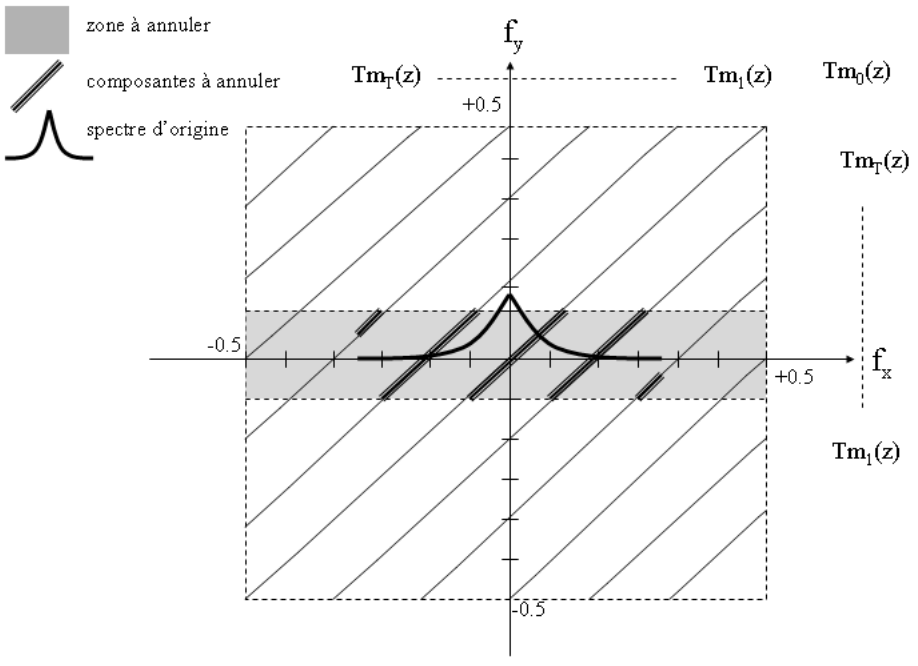

FIG. 2.32 – Annulation d'une bande fréquentielle de $\mathbf{v}$ pour un signal $\mathbf{u}$ à bande limitée

Application de ZI-LPTV au tatouage d'images

Grâce au parcours de Peano-Hilbert, il est possible de calculer le spectre 1D d'une image. Celui-ci décroît rapidement, la majeure partie de la puissance se concentrant dans les basses fréquences (cf. *fig.* 2.33). Cette caractérisation du spectre est commune à l'ensemble des images naturelles. Il est donc possible de construire un filtre LPTV indépendant de $\mathbf{x}$, ce qui ne serait pas le cas si l'on travaillait sur le spectre 2D d'une image particulière. On constate empiriquement que $99,5\%$ de la puissance du spectre d'une image après parcours de Peano est comprise dans la bande $[-\frac{1}{16}, \frac{1}{16}]$. La condition (2.16) est donc :

$$a(k,l) = 0 \quad \text{si}$$
$$(k,l) \in \{\frac{T}{2} - (\text{Int}(\frac{T}{16}) + 1), \ldots, \frac{T}{2} + (\text{Int}(\frac{T}{16}) + 1)\} \times \{\frac{T}{2} - (\text{Int}(\frac{T}{P}) + 1), \ldots, \frac{T}{2} + (\text{Int}(\frac{T}{P}) + 1)\}$$

Le fait que le filtre LPTV inverse crée un trou lorsque son entrée est un signal basse fréquence n'empêche pas le filtre LPTV direct d'étaler le spectre.

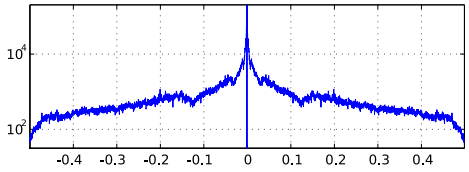

FIG. 2.33 – Spectre de l'image après parcours de Peano (Babouin)

Résultats d'étalement

La *fig.* 2.34 montre que le spectre de **m** peut être retrouvé après simple filtrage passe-bas. La *fig.* 2.35 montre que sur une image synthétique stationnaire à bande limitée dans $[-\frac{1}{16}, \frac{1}{16}]$, ZI-LPTV permet une annulation du spectre sur la bande visée (ici : $[-\frac{1}{256}, \frac{1}{256}]$). La *fig.* 2.36, montre que le spectre du message étalé est plus faible, mais pas nul, dans les basses fréquences qui sont prépondérantes dans l'image. Par contre, en sortie du LPTV inverse, les interférences de l'image sont presque nulles dans la partie du spectre correspondant à la bande du message. L'annulation n'est pas parfaite car la propriété de bande limitée n'est qu'une approximation. L'amélioration des performances de ZI-LPTV sur la version de base de mod-LPTV autorise l'emploi de la mise en forme NRZ.

Matrice A du filtre LPTV (resp. du filtre LPTV inverse), $T = 256$, $L = 1024$

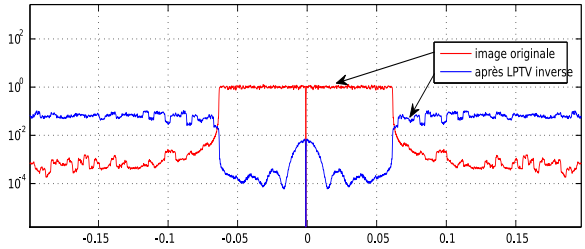

FIG. 2.34 – Spectre en sortie de ZI-LPTV inverse, image synthétique stationnaire (TEB=0) $T = 256$, $L = 4096$, DWR=15 dB

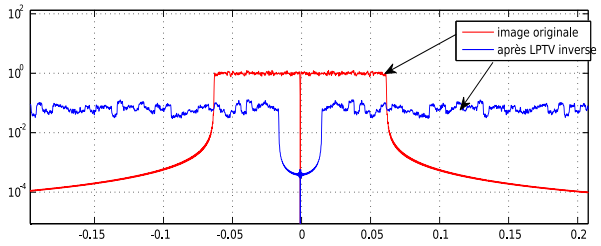

FIG. 2.35 – Effet de ZI-LPTV sur une image synthétique stationnaire à bande limitée

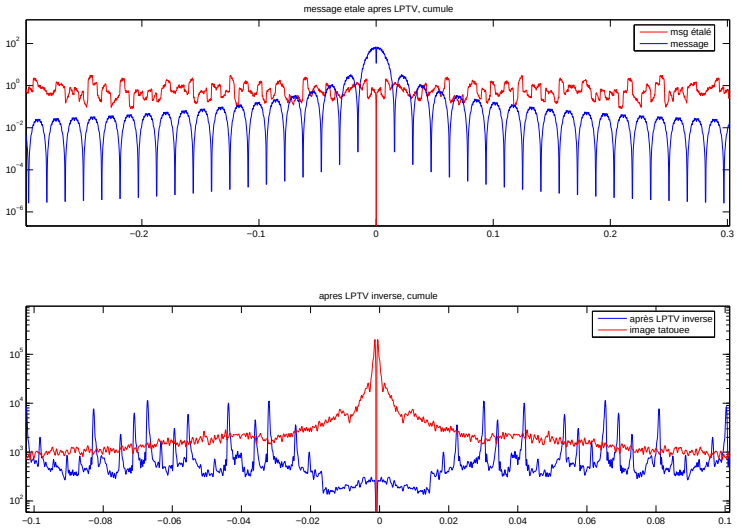

FIG. 2.36 – Influence spectrale de ZI-NRZ-LPTV et de son inverse, DWR=28 dB, $L = 4096$, $T = 512$

Comparaison avec DS et préfiltrage passe-haut avant démodulation

Le principe d'éliminer l'influence des composantes basses fréquences de $\mathbf{x}$ à la démodulation peut être appliqué à DS avec parcours de Peano et mise en forme NRZ. On applique alors un filtre passe-bas avant démodulation. Pour que la technique soit inversible (TEB nul en l'absence de bruit), $\mathbf{w}$ ne doit pas posséder de composantes fréquentielles dans la bande coupée. Cette technique n'est cependant pas utilisée pour DS, entre autres parce que la mise en forme NRZ est sensible au rognage. De plus, si

$\mathbf{w}$ ne modifie que les hautes et moyennes fréquences, il est moins robuste. ZI-LPTV présente l'avantage d'être inversible tout en générant un tatouage à spectre étalé sur toute la bande. ZI-LPTV offre une dispersion plus importante du support d'insertion si T est grande. Sur la *fig.* 2.37, on voit que ZI-DS, initialement plus performant que DS+W, est sensible au rognage. ZI-LPTV offre une meilleure robustesse.

Face aux réserves exprimées ci-dessus, les basses fréquences de l'image sont souvent éliminées par un filtre adaptatif de Wiener (cf. paragraphe 1.5.5), qui offre d'excellentes performances. ZI-LPTV sera combiné à un tel filtre dans le paragraphe 2.4.1.

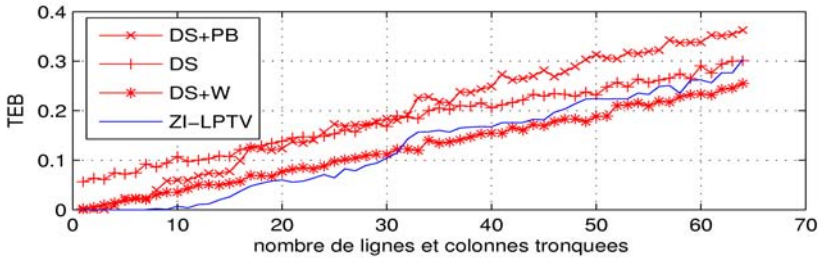

FIG. 2.37 – Robustesse de ZI-LPTV et de ZI-DS au rognage, $L = 512$, DWR=20 dB, $T = 512$, $N = 2^{16}$

2.3.4 Filtres LPTV et masque spectral

Objectif de mask-LPTV : la formulation modulateur d'un filtre LPTV permet d'imposer des contraintes sur le spectre de sortie à partir du spectre d'entrée. Appliquée au masquage perceptuel, cette propriété permet de construire facilement des tatouages situés dans les hautes ou moyennes fréquences, donc moins perceptibles. Cependant, un tel tatouage serait moins robuste à la compression ou au débruitage qu'un tatouage à spectre étalé. Un problème plus intéressant consiste à construire un tatouage dont le spectre a la même forme que celui du document original. On nommera cette technique **mask-LPTV**. Image et tatouage seront semblables spectralement, mais seront différents. Il a été observé heuristiquement que cette propriété, appelée "Contrainte du Spectre de Puissance" (PSC, *Power-Spectrum Condition*), augmente la robustesse du tatouage. En particulier, elle permet de maximiser la robustesse à l'attaque de débruitage par filtrage de Wiener (estimation MMSE du tatouage) [SEG01][SG02]. Pour la méthode DS, on doit opérer un filtrage passe-bas du tatouage après modulation pour qu'il respecte la PSC.

Implantation : dans un premier temps, on simplifie la synthèse du filtre en considérant des filtres modulateurs constants, comme dans mod-LPTV. Soit $\mathcal{P}_k$ la puissance de l'image sur la bande $[\frac{k}{T}, \frac{k+1}{T}]$. La contrainte imposée sur le filtre LPTV d'insertion est que pour un message à mise en forme NRZ, la puissance de $\mathbf{v}$ contenue dans la bande soit proportionnelle à celle de l'image. Avec la mise en forme répétition, on suppose le spectre de $\mathbf{b}$ plat. Au besoin, on applique préalablement une modulation par un code pseudo-aléatoire (ou une permutation, un PCC) pour blanchir le message de départ. La contrainte est alors

$$\sum_l u(k,l) = \mathcal{P}_k$$

On fixe donc la norme de la ligne de A correspondant à chaque zone spectrale, à partir par exemple d'une matrice orthonormale aléatoire. L'orthogonalité subsiste, mais pas la normalité. La matrice inverse n'est plus orthonormale. Si U et A sont constants par morceaux, alors V aussi. Dans un second temps, on pourrait donc prendre pour objectif une pondération linéaire par morceaux. Il faudrait introduire des retards dans A, et résoudre un système d'équations pour l'inversibilité ou pour la synthèse.

Application à l'image : l'implantation de mask-LPTV soulève deux problèmes majeurs pour un document image. D'une part, comme l'image est concentrée dans les basses fréquences, T doit être grande pour obtenir une approximation fine du spectre de $\mathbf{x}$, ce qui conduit à un grand coût calculatoire. D'autre part, si le spectre du tatouage se concentre dans les basses fréquences, le détecteur sera nécessairement plus sensible au bruit de transmission à travers l'image, qui est également de basse fréquence. Pour $T = 512$ et avec préfiltrage de Wiener (cf. paragraphe 1.5.5), le spectre épouse celui de l'image (cf. *fig.* 2.38) mais les résultats au décodage sont inférieurs à ceux de mod-LPTV à cause de la mise en forme NRZ (TEB=0.02 pour $L = 100$ et DWR=28 dB).

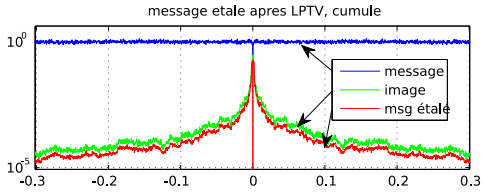

FIG. 2.38 – Exemple de masque spectral obtenu par un LPTV et une mise en forme répétition

2.3.5 Performances théoriques face au bruit additif blanc gaussien

Cas où l'hôte est blanc gaussien : lorsque l'hôte est blanc gaussien, toutes les techniques LL-LPTV, LL-NRZ-LPTV, mod-LPTV et mod-NRZ-LPTV ont des performances similaires à DS (cf. *fig.* 2.39). La mise en forme n'a pas d'impact sur les performances. Les performances expérimentales sont très proches des performances théoriques. ZI-LPTV présente des performances légèrement moins bonnes, du fait du filtrage passe-bas en réception. Par contre, lorsque 99% de la puissance de l'hôte est concentrée dans une bande limitée, ZI-LPTV fournit d'excellentes performances, très proches de ses performances théoriques (cf. *fig.* 2.40) : la contribution du document hôte est divisée par 100.

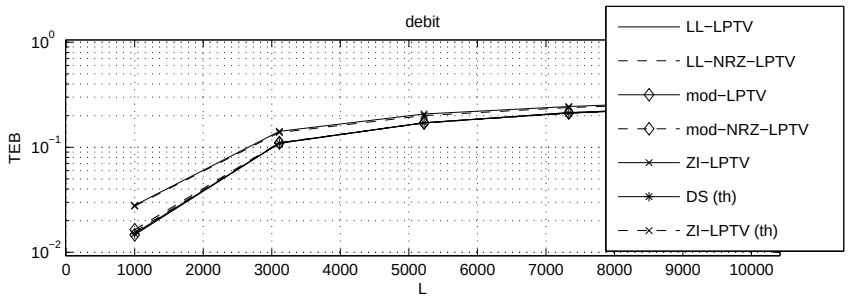

FIG. 2.39 – Hôte blanc gaussien : influence du débit, DWR=7 dB

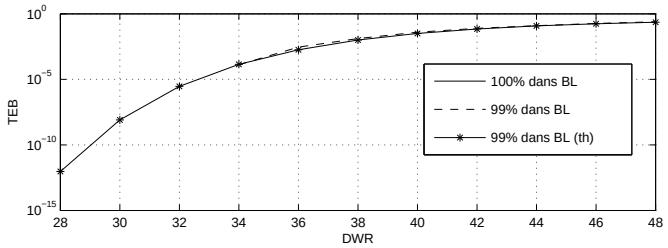

FIG. 2.40 – ZI-LPTV : image synthétique stationnaire à bande limitée, $L = 512$

Cas d'une image naturelle : la *fig.* 2.41 montre que la sensibilité des filtres LPTV au bruit d'une image est très variable selon L. Cette influence, liée à T, est difficile à expliquer théoriquement du fait de la non-stationnarité des images. On peut cependant noter que LL-LPTV et mod-LPTV ont toujours des performances supérieures à DS (à l'exception de mod-LPTV lorsque T est multiple de L : $L = 64$ ou $L = 128$), et souvent bien supérieures (cas de $L = 76, 96, 128, 152, 192, 320, 512$). LL-LPTV a des performances légèrement supérieures à mod-LPTV. ZI-LPTV est beaucoup moins sensible à L, preuve que cette sensibilité est due aux basses fréquences de l'image, et fournit les meilleures performances. Enfin, la mise en forme NRZ conduit à des interférences très importantes et difficiles à évaluer.

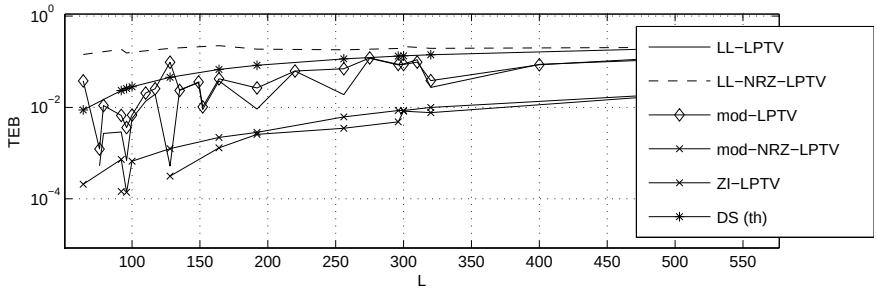

FIG. 2.41 – Influence de L sur les performances des filtres LPTV, DWR=28dB

2.3.6 Famille de filtres LPTV orthogonaux

LL-LPTV et mod-LPTV sont soumis à des interférences multi-utilisateurs lorsque $J > 1$ (cf. *fig.* 2.42). On peut montrer que si $J = 2$ et A_1, A_2 sont les matrices respectives des utilisateurs 1 et 2, les MAI valent

$$\sigma^2_{\text{MAI}} = P\Psi^2 ||A_1^{-1} A_2||^2$$

Les MAI diminuent avec L, quand la redondance augmente. Elles diminuent également lorsque T augmente : il y a une meilleure orthogonalité entre les matrices générées aléatoirement. LL-LPTV et mod-LPTV ne fournissent pas de garantie d'orthogonalité dans le cadre d'une application multi-utilisateurs. En effet, il est impossible de construire des matrices orthonormales et orthogonales entre elles. L'ensemble des matrices orthogonales est un groupe donc le produit de deux matrices orthogonales est orthogonal, et ne peut pas être proche d'une matrice nulle.

Une solution consiste à adapter l'algorithme mod-LPTV au cas multi-utilisateurs. Dans [Cha04], p.147, une condition suffisante pour construire une famille de filtres orthogonaux à partir de la représentation modulatrice est démontrée. Considérons un

filtre LPTV inversible de période T de filtres modulateurs $Tm_k^{(1)}(z), k = 1, \dots T$, dit "filtre générateur". Pour $j \in \{2, \dots, T\}$, on définit $T-1$ filtres LPTV par leurs filtres modulateurs $Tm_k^{(j)}(z)$ tels que :

$$Tm_k^{(j)}(z) = Tm_{\lfloor k-j \rfloor}^{(1)}(zW_T^{-j})$$

Il s'agit d'une permutation circulaire vers la droite d'ordre j des colonnes de la matrice modulateurs du filtre générateur. On peut montrer que chacun des T filtres LPTV est alors inversible. Considérons désormais que le signal $\mathbf{u}$ est à bande limitée : son spectre est inclus dans $[-\frac{b}{2}, \frac{b}{2}]$. Soient

$$\beta = \text{Int}(bT) + 1 \quad \text{et} \quad \gamma = \text{Int}(T/\beta)$$

Alors le sous-ensemble des γ filtres LPTV définis par les filtres modulateurs $Tm_k^{(j\beta)}(z)$ avec $j \in \{1, \dots, \gamma\}$ forme un système de filtres LPTV orthogonaux sur les signaux de bande inclue dans $[-\frac{b}{2}, \frac{b}{2}]$.

Dans [Cha04], cette technique de construction est appliquée à deux filtres LPTV générateurs : l'entrelaceur lignes/colonnes et le PCC linéaire, qui sont des filtres LPTV inversibles. Ici, nous proposons de l'appliquer à mod-LPTV. Nous utilisons la mise en forme NRZ pour le message, afin d'avoir un signal à bande limitée. Par exemple, si $T = 128$ et $L = 2048$, il y a au maximum $J = 64$ utilisateurs possibles. Cette technique, que nous appellerons **orth-LPTV**, a des performances équivalentes à mod-NRZ-LPTV lorsque $J = 1$, donc nettement moins bonnes que celles de mod-LPTV et LL-LPTV. Lorsque $J > 1$, l'orthogonalité est parfaite. pour orth-LPTV, qui présente alors de meilleures performances que LL-LPTV et mod-LPTV. Les performances de orth-LPTV décroissent lorsque J augmente à cause de la perte de puissance de chaque tatouage. Il serait possible de combiner ZI-LPTV avec orth-LPTV.

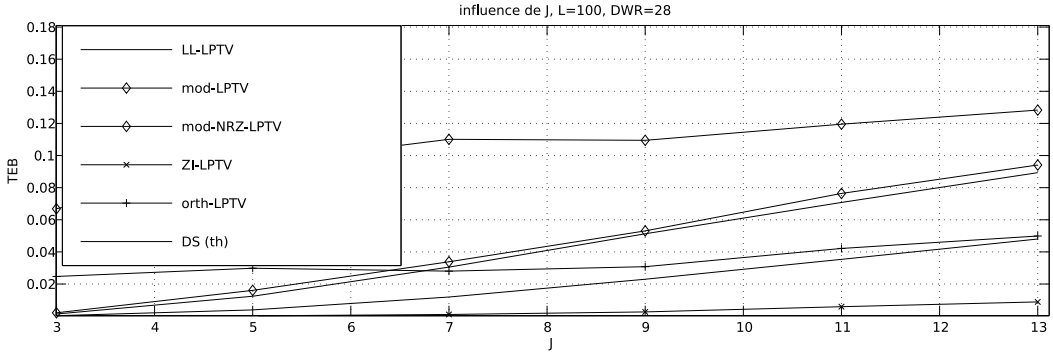

FIG. 2.42 – Interférences multi-utilisateurs, DWR=28 dB, $L = 100$

2.3.7 Etude perceptuelle : application à l'image

Absence de masque perceptuel : la *fig.* 2.43 montre des exemples de tatouages générés par les filtres LPTV. Aucun artefact (motifs...) n'est visible. Les distances de Watson, SSIM et Kullback-Leibler (cf. paragraphe 1.5.3) sont utilisées pour évaluer les différents filtres LPTV à DWR donné sur une base de 44 images [Cit]. Ici, DWR=28 dB et PSNR=43.5 dB. Selon le critère SSIM , les performances perceptuelles des filtres

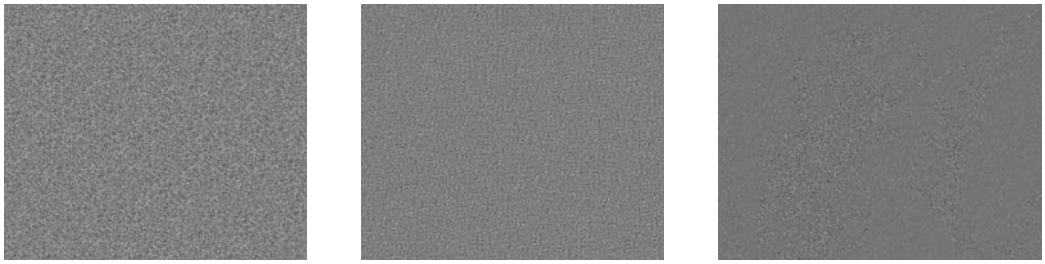

FIG. 2.43 – Exemples de tatouage : (gauche) LL-LPTV, (centre) ZI-LPTV, (droite) LL-LPTV+NVF

LPTV sont similaires à celle de DS. Par contre, le fait que le tatouage inséré ne soit pas blanc se répercute sur les critères de Watson et de Kullback-Leibler. Selon ces critères, les filtres LPTV doivent donc être combinés avec un masque perceptuel.

	SSIM	D_W	D_{KL}
DS	0.9870	365	1370
LL-LPTV	0.9887	1443	1778
LL-LPTVNRZ	0.9880	1187	1718
mod-LPTV	0.9887	1443	1733
mod-NRZ-LPTV	0.9880	1063	1448
ZI-LPTV	0.9881	1055	1456

Combinaison avec un masque perceptuel : le comportement des filtres LPTV combinés avec un masque perceptuel est similaire à celui de DS (cf. *fig.* 2.43). Les performances sont peu affectées par l'utilisation du masque (cf. *fig.* 2.44). Comme pour DS, la robustesse à certaines attaques peut même être améliorée par l'utilisation de masques. Le masque utilisant l'interpolation bilinéaire effectue une sorte de pré-blanchiment qui élimine une partie des interférences de l'image. Combinés à un masque perceptuel, les filtres LPTV présentent des performances similaires à DS selon les critères SSIM et Kullback. L'utilisation d'un masque reposant sur l'interpolation bilinéaire améliore également le critère de Watson.

	SSIM	D_W	D_{KL}
DS+NVF	0.9944	317	709
LL-LPTV+NVF	0.9953	1186	678
DS+masque bilin	0.9978	347	255
LL-LPTV+masque bilin	0.9978	580	229

En sortie du LPTV, le tatouage n'est pas à valeurs antipodales comme avec les techniques DS avec code antipodal et PCC, mais suit une distribution gaussienne (cf. *fig.* 2.45). Même sans l'emploi d'un masque, ce type de distribution du tatouage convient parfaitement à l'imperceptibilité (on la retrouve notamment avec la technique DS avec séquence **c** gaussienne).

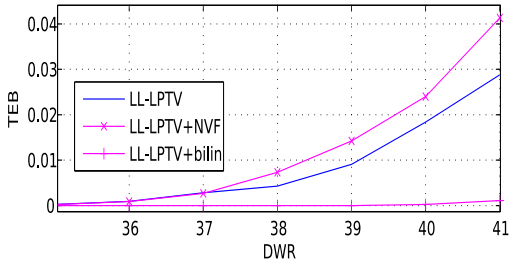

FIG. 2.44 – Performance selon le masque perceptuel en fonction de DWR, $L = 100$

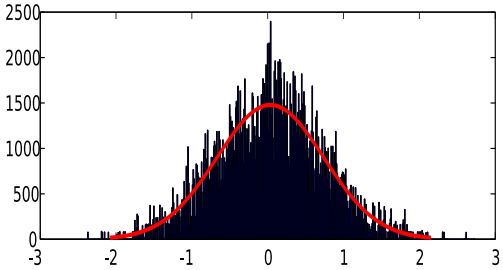

FIG. 2.45 – Histogramme d'un tatouage généré avec mod-LPTV

2.3.8 Etude de la robustesse : application à l'image

Les résultats expérimentaux sont présentés dans l'annexe A.2.1, nous en présentons un récapitulatif ci-dessous.

	PCC+Peano	LL-LPTV	mod-LPTV
dépendance à L		très dépendant	très dépendant
dépendance à T		faible si $T > 100$	faible
robustesse à $\mathbf{x}$	bonne : division de $\sigma^2_{\mathbf{x}}$ par 3	bonne : division de $\sigma^2_{\mathbf{x}}$ par 3	bonne : division de $\sigma^2_{\mathbf{x}}$ par $\frac{3}{2}$
tatouage multiple	similaire à DS	impossible	impossible
AWGN	similaires	similaires	similaires
débruitage	bonnes (selon L)	bonnes (selon L)	bonnes (selon L)
compression	très bonnes	très bonnes	très bonnes
	LL-NRZ-LPTV	mod-NRZ-LPTV	ZI-LPTV
dépendance à L	faible	faible	faible
dépendance à T	faible si $T > 100$	T choisi grand	T choisi grand
robustesse à $\mathbf{x}$	très mauvaise	très mauvaise	excellente : division de $\sigma^2_{\mathbf{x}}$ par 100
tatouage multiple	adaptation compliquée	adaptation orth-LPTV	adaptation possible
AWGN	similaires	similaires	similaires
débruitage			excellentes
compression			excellentes

2.4 Exploitation des propriétés statistiques d'une image

Dans cette partie, nous appliquons aux techniques utilisant les filtres LPTV les améliorations à la chaîne de tatouage DS présentées dans le paragraphe 1.5.5.

2.4.1 Domaine spatial : pré-blanchiment

Nous appliquons ici le même préfiltrage que pour DS (cf. paragraphe 1.5.5). Les simulations sont présentées dans l'annexe A.2.2. Pour l'ensemble des techniques étudiées, à l'exception de ZI-LPTV, le préfiltrage de Wiener correspond à une division de $\sigma^2_{\mathbf{x}}$ par 20. Cependant, les filtres LPTV perdent une grande partie de leurs spécificités : les performances sont peu dépendantes de L, donc plus proches de celles de DS+W.

Seul ZI-LPTV doit faire l'objet d'une adaptation. Lorsqu'on combine ZI-LPTV et préfiltrage de Wiener, on combine un filtre passe-haut dans le domaine spectral avec un passe-haut adaptatif dans le domaine spatial. En effet, en utilisant un préfiltrage de Wiener pour stationnariser l'image, les interférences sont plus importantes dans les moyennes fréquences (cf. *fig.* 2.46), même s'il reste des composantes dans les basses fréquences (cf. *fig.* 2.47). Le changement de l'allure du spectre impose une modification de ZI-LPTV+W. La partie du spectre de l'image annulée est alors élargie aux moyennes fréquences. Une modélisation plus fine du spectre (par exemple, identification de pics en $f = 0.2$) n'apporte pas d'amélioration significative des performances en pratique.

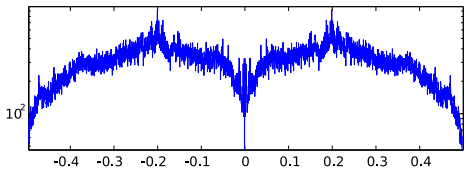

FIG. 2.46 – Spectre de l'image après parcours de Peano et préfiltrage de Wiener (Babouin)

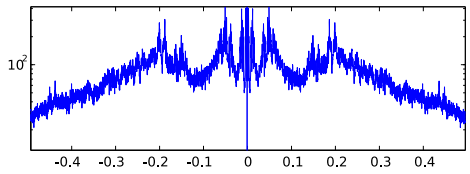

FIG. 2.47 – Spectre de l'image après parcours de Peano et préfiltrage de Wiener (Bateaux)

2.4.2 Domaine transformé : décodage optimal

L'adaptation du décodeur optimal de l'équation (1.17) aux PCC est :

$$\hat{d}_l^j = \frac{1}{P} \sum_{k \in \{(l-1)P+1,\ldots,lP\}} \frac{|f_j^{-1}(\mathbf{z})(k) + f_j^{-1}(\Psi)(k)|^{\hat{c}_{f_j^{-1}([1,\ldots,N])(k)}} - |f_j^{-1}(\mathbf{z})(k) - f_j^{-1}(\Psi)(k)|^{\hat{c}_{f_j^{-1}([1,\ldots,N])(k)}}}{\sigma_{f_j^{-1}([1,\ldots,N])(k)}^{\hat{c}_{f_j^{-1}([1,\ldots,N])(k)}}}$$

Pour les filtres LPTV, nous exprimons ici la statistique suffisante à partir de $t'_k = t'_k - w_k$, où **w** est généré à partir d'un LPTV. Contrairement aux cas DS et PCC, il n'y a pas indépendance entre les échantillons $\{w_k\}$ (on a opéré un filtrage avec mémoire). Il est donc impossible de calculer de façon simple une statistique suffisante pour $\hat{m}_l$, estimateur de m_l. Une stratégie sous-optimale consiste à obtenir une première estimation $\hat{\mathbf{m}}^0$ de **m** par un décodage classique. Puis le test sur le bit l est réalisé entre les deux hypothèses :
$\mathrm{H}_0 : m_l = -1, m_{l'} = m^0_{l'} \quad \forall l' \neq l$
$\mathrm{H}_1 : m_l = 1, m_{l'} = m^0_{l'} \quad \forall l' \neq l.$

$$\hat{d}_l = \frac{1}{P} \sum_{k \in \mathcal{S}_l} \frac{|t'_k - \psi_k \mathcal{F}^{\mathrm{LPTV}}(M|\mathrm{H}_1)|^{\hat{c}_k} - |t'_k - \psi_k \mathcal{F}^{\mathrm{LPTV}}(M|\mathrm{H}_0)|^{\hat{c}_k}}{\sigma_k^{\hat{c}_k}}$$

est une statistique suffisante. L'estimateur $\hat{\mathbf{m}}^1$ de **m** ainsi construit est pour chaque bit l, $\hat{m}^1_l = \text{signe}(\hat{d}_l)$. D'autres itérations peuvent éventuellement être effectuées, au détriment du coût calculatoire.

L'utilisation du décodeur optimal apporte une amélioration de la robustesse au bruit de l'hôte similaire à celle de DS (cf. Figs. 2.48 et 2.49). Dans le domaine de la DCT, les PCC et filtres LPTV perdent les particularités constatées dans le domaine spatial des images naturelles.

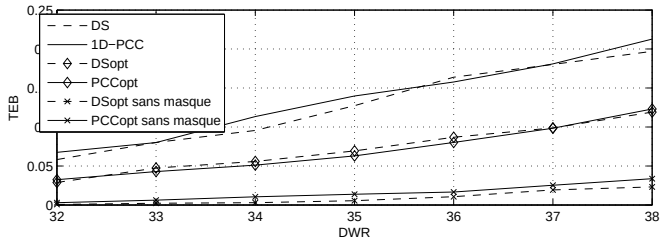

FIG. 2.48 – Décodeur optimal dans le domaine de la DCT : robustesse au bruit de l'hôte, $L = 100$

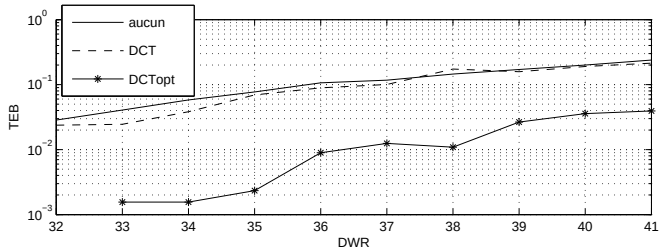

FIG. 2.49 – Insertion dans le domaine transformé : influence de DWR, LL-LPTV, L=60

2.5 Tatouage informé et filtres LPTV

L'étalement de spectre intervient dans deux types d'algorithmes pratiques de tatouage informé : l'étalement de spectre amélioré et les techniques quantificatives. Il est donc possible d'insérer les PCC et filtres LPTV proposés dans le chapitre précédent dans la chaîne de tatouage informé, à la place de la modulation classique par un code.

2.5.1 Étalement de spectre amélioré

Les simulations sont présentées dans l'annexe A.2.3. Dans le domaine spatial, la diminution de l'influence de $\mathbf{x}$ constatée pour PCC+Peano, LL-LPTV, mod-LPTV et ZI-LPTV se répercute sur l'étalement de spectre amélioré, dans les mêmes proportions que pour le tatouage à insertion aveugle. Par contre, les techniques d'étalement de spectre amélioré combinées au préfiltrage de Wiener éliminent une grande partie de la spécificité des filtres LPTV, à l'exception de ZI-LPTV.

Changements d'horloge améliorés (IPCC)

L'étalement de spectre amélioré linéaire défini dans l'équation 1.8 est étendu à 1D-PCC par :

$$\mathbf{w}' = \psi f(\alpha \mathbf{b}^j - \lambda \nu^j)$$

où $\forall l \in \{1,..,L\}, p \in \{1,..,P\}$,

$$\nu_{l+(p-1)L} = \frac{1}{\psi P} \sum_{p'=1}^{P} \left(f^{-1}(x_{l+(p'-1)L}) - \mu(\mathbf{x}) \right)$$

et dans le cas des 2D-PCC,

$$\mathbf{w}' = \psi f^1 \circ f^2(\alpha \mathbf{b} - \lambda \nu)$$

$$\text{où} \quad \nu_{\lceil l+(p-1)L \rceil, \lfloor l+(p-1)L \rfloor} = \frac{1}{P} \sum_{p=1}^{P} ((f^2)^{-1} \circ (f^1)^{-1}(\mathbf{z}))(\lceil l+(p-1)L \rceil, \lfloor l+(p-1)L \rfloor)$$

$\forall l \in \{1,..,L\}, p \in \{1,..,P\}$, avec $l+(p-1)L = N_1 \lceil l+(p-1)L \rceil + \lfloor l+(p-1)L \rfloor$.

Le décodage est identique à 1D-PCC, 2D-PCC. Après décodage :

$$\begin{aligned} \hat{d}_l &= \frac{1}{P} \sum_{p=1}^{P} (f^{-1}(\mathbf{z}))(l+(p-1)L) \\ &= \alpha \psi m_l + \frac{1-\lambda\psi}{P} \sum_{p=1}^{P} f^{-1}(\mathbf{x})(l+(p-1)L) + \lambda\psi\mu(\mathbf{x}) + \frac{1}{P} \sum_{p=1}^{P} (f^{-1}(\mathbf{n}))(l+(p-1)L) \end{aligned}$$

On a désormais $E[\hat{d}] = \alpha\psi\mathbf{m} + \mu(\mathbf{y})$ et $\text{Var}(\hat{d}) = ((1-\lambda)^2\sigma_{\mathbf{y}}^2 + \sigma_{\mathbf{n}}^2)/\psi^2 P$.

Pour 1D-IPCC, le préfiltrage à l'étalement (IPCC+W) correspond à

$$\nu^j_{l+(p-1)L} = \frac{1}{\psi P} \sum_{p'=1}^{P} \left(f_j^{-1}(\mathbf{x} - \hat{\mathbf{x}})(l+(p'-1)L) - \mu(\mathbf{x}) \right)$$

Filtres LPTV améliorés (ILPTV)

Le tatouage par filtres LPTV améliorés consiste à insérer le tatouage préfiltré **w'** :

$$\mathbf{w}' = \psi \mathcal{F}^{\text{LPTV}}(\alpha \mathbf{b} - \lambda \nu)$$

où $\forall l \in \{1,..,L\}, p \in \{1,..,P\}$,

$$\nu_{|\mathcal{S}_l} = \frac{1}{\psi P} \sum_{p \in \mathcal{S}_l} \left((\mathcal{F}^{\text{LPTV}^{-1}}(\mathbf{x}))(\mathcal{S}_l) \right)$$

Les calculs effectués pour LISS sont encore valables car le filtre est linéaire. Pour calculer α et λ, une adaptation est nécessaire pour les filtres LPTV. Rappelons que λ contrôle l'erreur introduite par le document support lors du décodage. Ainsi, pour conserver la même distorsion $\sigma_{\mathbf{c}}^2$ que dans DS, l'équation (1.9) est toujours valable. Par contre, l'équation (1.10) devient :

$$\text{TEB} = Q\left(\sqrt{\frac{P\sigma_{\mathbf{w}}^2 - \lambda^2\sigma_{\mathbf{x}}^2}{(1-\lambda)^2\sigma_{\mathbf{x},LPTV,L}^2 + \sigma_{\mathbf{n}}^2}}\right).$$

On remarque que cette fois-ci, deux variances liées à l'hôte interviennent dans le calcul : $\sigma_{\mathbf{x}}^2$ et $\sigma_{\mathbf{x},LPTV(L)}^2$. Au lieu de l'équation (1.11), Le TEB est optimal pour

$$\lambda_{\text{opt}} = \frac{1}{2}\left((1 + \frac{\sigma_{\mathbf{n}}^2}{\sigma_{\mathbf{x},LPTV(L)}^2} + \frac{P\sigma_{\mathbf{w}}^2}{\sigma_{\mathbf{x}}^2}) - \sqrt{\left(1 + \frac{\sigma_{\mathbf{n}}^2}{\sigma_{\mathbf{x},LPTV(L)}^2} + \frac{P\sigma_{\mathbf{w}}^2}{\sigma_{\mathbf{x}}^2}\right)^2 - 4\frac{P\sigma_{\mathbf{w}}^2}{\sigma_{\mathbf{x}}^2}}\right)$$

En l'absence d'une estimation empirique fine de $\sigma_{\mathbf{x},LPTV}^2(L)$, on pourrait utiliser une approximation par $\sigma_{\mathbf{x},PCC,Peano}^2$ (cf. paragraphe 2.2.2). ZI-LPTV conserve son intérêt car en réduisant l'influence de l'hôte, on réduit aussi le bruit à compenser et plus de puissance est accordée au tatouage.

2.5.2 Méthodes quantificatives fondées sur les filtres LPTV

Les techniques ST-SCS, STDM et QP (cf. paragraphe 1.3.2) combinent étalement de spectre et tatouage quantificatif. Il est possible de s'en inspirer pour construire des méthodes quantificatives avec redondance par étalement fondée sur les PCC et les filtres LPTV. Nous appellerons ces méthodes PCC-QP, PCC-SCS et PCC-DM, ainsi que LPTV-QP, LPTV-SCS et LPTV-DM.

Utilisation des PCC

Soit P_s la taille du code d'étalement. Pour la technique QP, P_s est ici petit devant T (par exemple, $P_s = 20$). Pour ST-SCS et STDM, $P_s = P$ est grand. A la différence de STDM, ST-SCS inclut une distorsion des compensations. On découpe l'image en N/P_s sous-ensembles $\mathcal{S}_k^{P_s}$, eux-mêmes regroupés en L ensembles $\mathcal{S}_k$ correspondant à chaque pixel. Dans PCC-QP, T restera grand pour des raisons d'étalement et de sécurité. P_s n'influera donc que sur la taille du support de la moyenne effectuée à l'étalement. Avant permutation, les supports $\mathcal{S}_k^{P_s}$ ne sont pas choisis aléatoirement mais sont issus d'une mise en forme répétition :

$$\begin{aligned} x_{l,k} &= \frac{1}{P_s}\sum_{p=1}^{P_s} f_j(x_{l+kP_s+(p-1)L}) - \mu(\mathbf{x}) \\ &= Q_{\Delta,\tau_k}(x_{l,k}) + d_{l,k} - \tau_k \quad \longrightarrow \quad y_{k,l} = Q_{\Delta,\tau_k}(x_{l,k}) \end{aligned}$$

avec $k \in \{0,..,N/P_s - 1\}$, $l \in \{1,..,L\}$. Ce schéma correspond donc à

$$y_{l+kP_s+(p-1)L} = x_{l+kP_s+(p-1)L} - (d_{l,k} - \tau_k) \quad \forall p \in \{1,..,P_s\} \ .$$

L'intensité des échantillons de chaque sous-ensemble $\mathcal{S}_k^{P_s}$ est donc augmentée ou diminuée d'une valeur $d_{l,k} = y_{k,l} - x_{l,k}$ dépendant de $\mathcal{S}_k^{P_s}$.

La *fig.* 2.50 montre que PCC-SCS, du fait de la mise en forme PCC et de la fonction de projection inverse, peut générer des effets de bloc si L est court. Cet inconvénient n'apparaît pas avec ST-SCS et mise en forme aléatoire, grâce au code d'étalement.

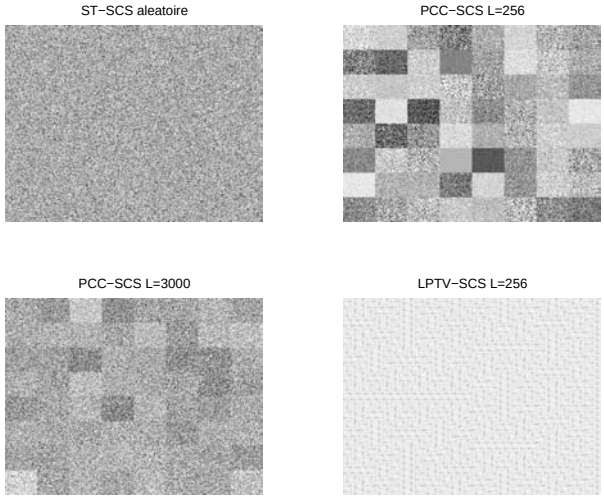

FIG. 2.50 – Exemples de tatouages générés par ST-SCS, PCC-SCS et LPTV-SCS

Utilisation des filtres LPTV

Nous proposons ici d'utiliser les filtres LPTV comme fonction d'étalement. On quantifie alors, selon la mise en forme :

$$\begin{aligned} x_{l,k} &= \frac{1}{P_s} \sum_{n \in \mathcal{S}_k^{P_s}} \mathcal{F}^{\mathrm{LPTV}^{-1}}(\mathbf{x})(n) \\ &= Q_{\Delta,\tau_k}(x_{l,k}) + q_{l,k} - \tau_k \quad \longrightarrow \quad y_{k,l} = Q_{\Delta,\tau_k}(x_{l,k}) \end{aligned}$$

avec $k \in \{0, .., N/P_s - 1\}$, $l \in \{1, .., L\}$. Soit $\mathbf{q}$ le vecteur composé de l'erreur de quantification "mise en forme" : $q_n = q_{l,k} - \tau_n \forall n \in \mathcal{S}_k^{P_s}$. On insère ensuite grâce à la formule de reconstruction (*unprojection*) suivante :

$$\mathbf{y}_k = \mathbf{x}_k - \mathcal{F}^{\mathrm{LPTV}}(\mathbf{q})(k) \quad \forall k \in \mathcal{S}_k^{P_s} \quad .$$

Les performances théoriques de LL-LPTV-SCS et mod-LPTV-SCS en présence de bruit AWGN sont équivalentes.

Le fait de quantifier la moyenne de P_s échantillons en sortie du filtre LPTV est inspiré du décodeur du tatouage par LPTV classique. Si $P_s = P$, LPTV-SCS est une technique de redondance par étalement inspirée de ST-SCS. Si $P_s = 1$, LPTV-SCS se ramène à un SCS simple dans un "domaine transformé". Si $1 < P_s < P$, LPTV-SCS est une technique de quantification de la projection sur un sous-espace qu'on peut rapprocher de SSP (cf. partie 1.3.2). Dans ce dernier cas, il subsiste une redondance par répétition de facteur P/P_s et une extension à la quantification vectorielle est possible. La *fig.* 2.50 montre que grâce au filtrage, aucun effet de bloc n'est présent dans le tatouage.

2.5.3 Etude de la robustesse : application à l'image

En l'absence d'attaque, la méthode quantificative supprime totalement le bruit apporté par l'hôte, les LPTV n'ont donc *a priori* pas d'intérêt par rapport à l'étalement DS. Par contre, des attaques comme le débruitage, la compression ou une attaque géométrique suivie de resynchronisation introduisent un bruit dépendant de l'image. Une étude expérimentale est nécessaire pour connaître l'intérêt de PCC-QP et LPTV-QP.

La *fig.* 2.51 montre l'amélioration des techniques avec redondance par étalement par rapport à la redondance par répétition. La robustesse des techniques à redondance par répétition est similaire, LL-LPTV-SCS avec $P_s = 1$ étant légèrement meilleure. Parmi les techniques à redondance par étalement, PCC-SCS et LL-LPTV-SCS sont moins robustes. mod-LPTV-SCS apporte une robustesse similaire à ST-SCS. La mise en forme (NRZ ou répétition) du message initial n'influe pas sur les performances. La *fig.* 2.52 montre que mod-LPTV-SCS est légèrement plus robuste à la compression JPEG que ST-SCS. Les autres techniques de type LPTV-SCS sont moins robustes. La combinaison de PCC-SCS ou ST-SCS avec un parcours de Peano n'a pas d'impact sur la robustesse. La *fig.* 2.53 montre que PCC-SCS est plus robuste à un bruit d'interpolation que les autres techniques. Les techniques LPTV-SCS sont moins robustes.

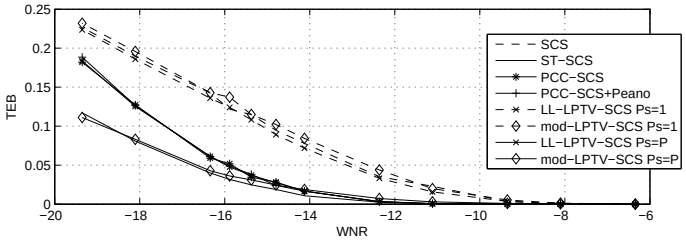

FIG. 2.51 – Robustesse au bruit AWGN, $L = 1300$, DWR=35 dB

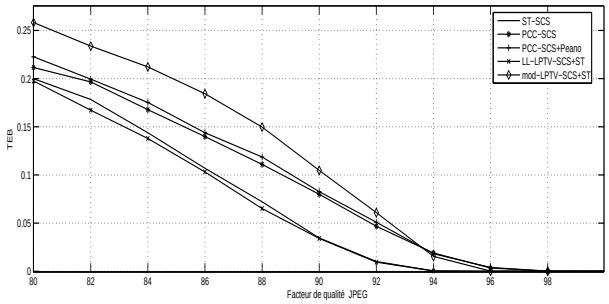

FIG. 2.52 – Robustesse à la compression JPEG, $L = 1300$, DWR=35 dB

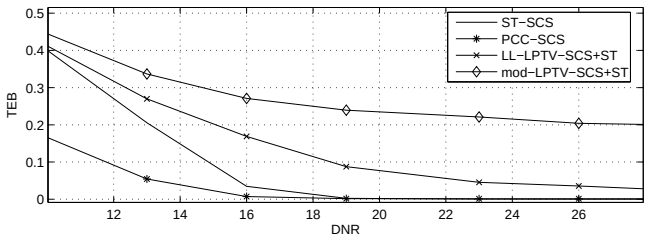

FIG. 2.53 – Robustesse au bruit d'interpolation, $L = 128$, DWR=28 dB

2.6 Sécurité des filtres LPTV

Dans un premier temps, on étudie dans ce paragraphe la sécurité des PCC, traités comme un cas particulier de technique DS. On propose une amélioration aux attaques pratiques existantes pour DS dans le cas où l'étalement se fait à base de PCC. La sécurité des techniques utilisant les filtres LPTV est ensuite étudiée.

2.6.1 Sécurité des PCC

En l'absence de code pseudo-aléatoire, la sécurité des PCC repose uniquement sur le secret de la mise en forme. Cette "modulation de la position des impulsions" [PFCTPPG06] est également utilisée pour la sécurité de techniques telles que le *patchwork* (cf. partie 1.2.7).

Entropie de la clé

Pour l'algorithme DS, il existe 2^N différentes clés binaires équiprobables de taille N, donc $\mathrm{H}(K) = N$. Pour les PCC, il existe $T!$ différentes permutations q de période T, donc on a pour 1D-PCC :

$$\mathrm{H}(q) = \log_2(T_{\mathrm{1D}}!)$$

et pour 2D-PCC :

$$\mathrm{H}(q^1, q^2) = 2\log_2(T_{\mathrm{2D}}!)$$

Comme $T_{\mathrm{2D}} << T_{\mathrm{1D}}$, 2D-PCC n'offre pas de bonnes garanties de ce point de vue. Pour ce critère précis, 1D-PCC est plus sûr que DS dès que $T > N/8$. Cependant, si les échantillons correspondant à chaque bit sont répartis aléatoirement et secrètement pour DS (mise en forme aléatoire "secrète"), cet avantage disparaît. $\mathbf{k}$ étant ici discrète, les réserves sur l'utilisation de l'approche de Shannon sont levées.

Approche de Fisher

On peut considérer les PCC sous une approche matricielle :

$$\mathbf{w} = \psi \sum_{j=1}^{J} \sum_{l=1}^{L} m_l^j f_j(\delta'_l) = \psi \sum_{j=1}^{J} \mathbf{m}^j f_j(\delta') \ ,$$

où $\delta'_l = 1$ si $k = l + pL$ et $\delta'_l = 0$ sinon. On retrouve alors un schéma de tatouage de la forme de celui étudié dans [CFF05]. Soient $N_c = LJ$, $\mathbf{u_l} = \frac{f_j(\delta'_l)}{\sqrt{P}}$, $\gamma = \psi\sqrt{LJ}\sqrt{P} = \psi\sqrt{N}\sqrt{J}$. Supposons que les porteuses soient de supports disjoints. Il y a donc une porteuse par bit. La formulation de schéma est alors :

$$\mathbf{w} = \frac{\gamma}{\sqrt{N_c}} \sum_{l=1}^{N_c} m_l \mathbf{u_l} = \frac{\gamma}{\sqrt{N_c}} \mathbf{m} U$$

Soient $\mathcal{W} = [\mathbf{w}^1, \ldots, \mathbf{w}^{N_o}]$, $\mathcal{X} = [\mathbf{x}^1, \ldots, \mathbf{x}^{N_o}]$, $\mathcal{Y} = [\mathbf{y}^1, \ldots, \mathbf{y}^{N_o}]$, $\mathcal{M} = [\mathbf{m}^1, \ldots, \mathbf{m}^{N_o}]$. On a : $\mathcal{W} = \frac{\gamma}{\sqrt{N_c}} U\mathcal{M}$. Les résultats de [CFF05] sont alors applicables (cf. paragraphe 1.4). Notamment, la fuite d'information sur $\mathbf{k}$ est linéaire en fonction du nombre d'observations, et le niveau de sécurité dépend de DWR. Les algorithmes pratiques d'attaque sur la sécurité utilisent une estimation ML ou des techniques de séparation de sources.

Lien entre DS et les PCC

Si $J = 1$, on peut réécrire les PCC sous forme de L porteuses disjointes dans $\{0,1\}^N$, donc sous la forme DS. On peut également considérer que les PCC sont une méthode DS sans clé secrète ($\mathbf{c} = 1_N$), la sécurité et l'étalement étant fondés sur l'emplacement des points (on parlerait alors de "mise en forme PCC"). Les PCC sont donc des codes DS discrets et structurés. Les contraintes suivantes s'appliquent. On utilise L porteuses disjointes ayant chacune P valeurs $+1$ et $N-P$ valeurs 0. Chacune des porteuses est de support $\mathcal{S}_l$. Enfin, les valeurs $+1$ sont réparties en "blocs" de taille T. Sous la forme matricielle (2.6.1) : $U = \{u_{l,k}, l = 1 \dots L\}$ et $\sum_{k=1}^{N} u_{l,k} = P$, $\sum_{l=1}^{L} u_{l,k} = 1$.

Le niveau théorique de sécurité des PCC est identique à celui de DS. La différence entre les deux techniques réside dans l'entropie de la clé. Cependant, le fait que les PCC soient considérés comme un sous-ensemble de codes DS discrets et structurés permet d'améliorer les algorithmes d'attaque sur la sécurité existants.

Algorithme pratique d'attaque sur les PCC

Nous proposons d'améliorer les estimateurs de [CFF05] pour les PCC par un module d'estimation de "motif" $N \times L$, utilisant une décision souple itérative. Ce système est également appliquable à DS avec codes binaires antipodaux de type CDMA. La reconnaissance d'un code structuré à partir de décisions souples bruitées est un problème de décodage itératif proche du décodage canal. Nous proposons dans cette section de construire un décodeur inspiré du décodage SISO (*Soft Input-Soft Output*) d'un code convolutif [Bou98]. Ce décodeur est sous-optimal au sens du maximum de vraisemblance (ML) mais ses performances tendent vers celles d'un décodeur ML lorsque le SNR augmente. Le coût calculatoire de cet algorithme de reconnaissance de code est très important, c'est pourquoi nous n'en donnons que le principe.

Soit $\hat{U} = U + N$ la matrice issue de l'estimation de U par une attaque sur la sécurité (cf. paragraphe 1.4.3). La première étape du décodeur est d'estimer la variance du bruit d'estimation. Puis on calcule indépendamment les résultats du décodage $u_{.,l}$ et $u_{k,.}$ respectivement sur les N lignes et les L colonnes $u_{k,l}$ en fonction des mots de code possibles. A chaque itération, on utilise le décodage des lignes pour améliorer celui des colonnes et vice-versa (cf. *fig.* 2.54).

La probabilité *a priori* d'un code u_k est [Bou98] :

$$\begin{aligned}
\mathrm{p}(U_k = u_k|\hat{U}) &= \sum_{U:\text{mot de code possible tel que } U_k = u_k} \mathrm{p}(U|\hat{U}) \\
&= \sum_{U|U_k=u_k} \frac{\mathrm{p}(\hat{U}|U)\mathrm{p}(U)}{\mathrm{p}(\hat{U})} \\
&= \frac{1}{\mathrm{p}(\hat{U})} \sum_{U|U_k=u_k} \mathrm{p}(\hat{U}|U)\mathrm{p}(U) \propto \sum_{U|U_k=u_k} \mathrm{p}(\hat{U}|U)\mathrm{p}(U)
\end{aligned}$$

$$\begin{aligned}
\mathrm{p}(U_k = u_k|\hat{U}) &\propto \sum_{U|U_k=u_k} \prod_{l=1}^{L} \mathrm{p}(\hat{U}_l|U_l)\mathrm{p}(U_l) \\
&\propto \mathrm{p}(\hat{U}_k = y_k|U_k = u_k)\mathrm{p}(U_l) \sum_{U|U_k=u_k} \prod_{l=1, l\neq k}^{L} \mathrm{p}(\hat{U}_l|U_l)\mathrm{p}(U_l)
\end{aligned}$$

En passant en logarithmique et en rapport 0/1 ($u_k \in \{0,1\}$), on définit les rapport des log-vraisemblances (LLR) suivants :

$$LLR(U_k|\hat{U}) = \ln \frac{\mathrm{p}(U_k=1|\hat{U})}{\mathrm{p}(U_k=0|\hat{U})} = LLR(U_k) + LLR(\hat{U}_k|U_k) + \ln \frac{\sum_{U|U_k=1} \prod_{l=1,l\neq k}^{L} \mathrm{p}(\hat{U}_l|U_l)\mathrm{p}(U_l)}{\sum_{U|U_k=1} \prod_{l=1,l\neq k}^{L} \mathrm{p}(\hat{U}_l|U_l)\mathrm{p}(U_l)}$$

Les probabilités *a priori* calculées par un décodeur SISO sont appelées "informations extrinsèques". Pour implanter le décodeur, on doit donc calculer les valeurs suivantes en fonction de la structure des codes à reconnaître : $LLR(U_{k,.}), LLR(U_{.,l}), LLR(U_{k,.} + R_{k,.}|U_{k,.})$ et $\sum_{U_{.,.}|U_{k,.}=1} \prod_{l=1,l\neq k}^{L} \mathrm{p}(\hat{U}_{l,.}|U_{l,.})\mathrm{p}(U_{l,.})$. A chaque itération t, le décodeur itératif calcule le $LLR(U_k|\hat{U})^{(t)}$ à l'aide des observations ($LLR(\hat{U}_{k,l}|U_{k,l})$) et d'une information *a priori* qui est le LLR extrinsèque issu du décodage sur les lignes ou les colonnes. La décision à l'itération t consiste à comparer $LLR(U_k|\hat{U})^{(t)}$ à un seuil , par exemple 0.

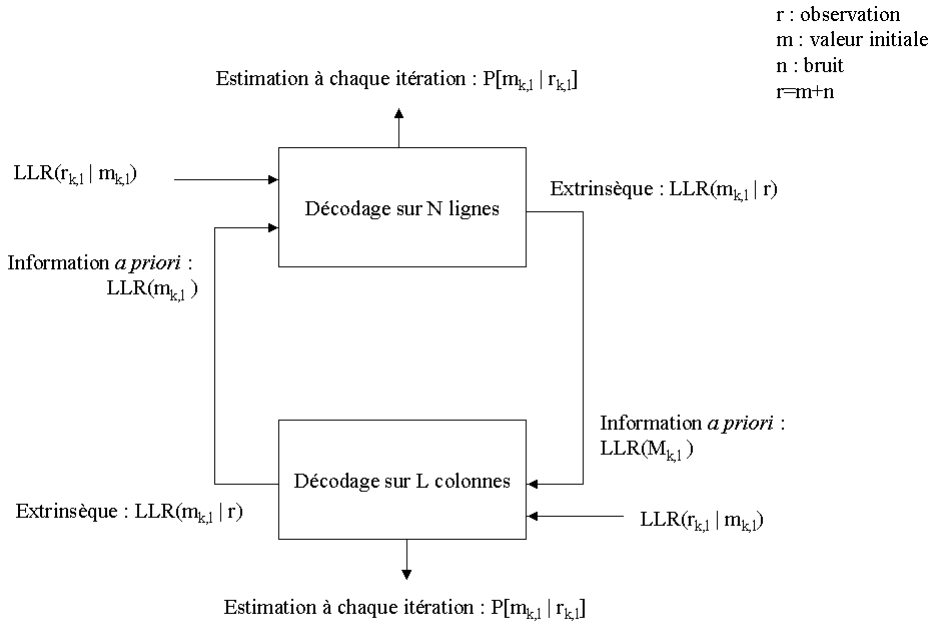

FIG. 2.54 – Schéma de l'estimateur de code PCC itératif

2.6.2 Sécurité des filtres LPTV

Dans le domaine spatial, l'existence du filtre complique grandement l'attaque sur la sécurité par rapport aux cas DS et PCC. Une analyse spectrale serait également possible, puisque le spectre n'est pas plat. Elle donnerait des indications sur le filtre sans révéler la clé elle-même. Nous montrons que les implantations simples des filtres LPTV que nous avons utilisées peuvent se ramener au cas de la sécurité de DS dans un domaine transformé (cf. paragraphe 1.4.2).

En effet, dans le cas LL-LPTV, le pirate peut se placer dans le domaine des composantes polyphases, où le tatouage est le résultat de $Y^{n_o} = UX^{n_o}$ pour $n_o \in$

$\{1, \ldots, N_o\}$. Dans les cas mod-LPTV, orth-LPTV, ZI-LPTV, du fait de l'utilisation de constantes, on peut se ramener à la résolution d'un système linéaire dans le domaine du bispectre. Il s'agit donc de la résolution d'un système linéaire (bruité dans le cas KMA), ou bien d'un problème de séparation de sources dans le cas WOA. La différence avec DS réside dans la taille de la matrice (TxT au lieu de NxJ) : on se ramène au cas DS avec une clé de taille T et T porteuses. Le nombre d'observations est multiplié par N/T, le niveau de sécurité est donc plus faible. En présence de retards dans LL-LPTV ou mod-LPTV, la résolution du système serait plus compliquée. Les attaques pratiques sur la sécurité des filtres LPTV utiliseraient donc les estimateurs ML, ICA et PCA classiques (cf. paragraphe 1.4.3). Les codes LPTV sont moins structurés que les codes PCC : seule l'orthonormalité réduit l'espace de recherche.

2.7 Conclusion : des techniques d'étalement de spectre alternatives

Les PCC : une alternative simple et efficace à DS

Les communications par étalement de spectre du type DS-CDMA ont inspiré de nombreuses techniques de tatouage. Les permutations aléatoires ont quant à elles été utilisées à différents niveaux du processus de tatouage. Cependant, aucune comparaison n'avait été effectuée jusqu'à présent entre la modulation DS et une technique d'étalement de spectre alternative utilisant les permutations aléatoires. Cette étude proposait d'utiliser les permutations aléatoires dans le cadre théorique plus général des PCC. Un algorithme de tatouage par étalement de spectre PCC, avec des permutations aléatoires 1D ou 2D, a été proposé et comparé à l'algorithme DS classique dans le domaine spatial et celui des blocs 8x8 de la DCT. Les PCC sont également adaptés au tatouage multiplicatif (cf. annexe B.2). Pour chacun de ces algorithmes, les performances théoriques ont été calculées. Les performances de DS et PCC sont globalement similaires et sont conformes à celles attendues pour des techniques d'étalement de spectre : une grande robustesse au bruit additif et multiplicatif mais une vulnérabilité aux attaques désynchronisantes. Une insertion dans le domaine de la DCT permet d'améliorer l'imperceptibilité et la résistance à la compression JPEG. Cette étude pourrait être étendue à d'autres domaines transformés ou masques perceptuels, ou à des supports moins limités en taille tels que l'audio ou la vidéo, où la redondance serait plus grande et la périodicité mieux exploitée.

Nous avons montré que la sécurité des PCC repose sur la modulation des positions des échantillons. De ce fait, elle est moindre que celle des techniques DS à mise en forme aléatoire secrète. Nous avons proposé, par analogie avec le décodage canal itératif, un module de reconnaissance des codes PCC destiné à améliorer contre les PCC les algorithmes pratiques d'attaque sur la sécurité existants.

Au final, même si dans le cas général les PCC utilisant les permutations aléatoires n'apportent pas d'amélioration nette des performances, ils permettent d'obtenir d'aussi bons résultats que l'étalement classique tout en étant plus simples de concept et d'implantation. On peut donc conseiller l'emploi des PCC pour les systèmes où les ressources sont critiques, comme la protection de copie des DVD. En effet, s'il n'est pas possible dans une application d'effectuer de pré-blanchiment au décodage ou de transformation de domaine pour des raisons calculatoires, les PCC offrent de meilleures performances, et utilisent des clés secrètes courtes.

Importance d'un parcours d'image pour le tatouage dans le domaine spatial

L'étude précédente a mis en évidence une amélioration de la robustesse à l'image par les PCC avec mise en forme répétition dans le domaine spatial. Dans le domaine spatial, 2D-PCC serait préféré à 1D-PCC pour son coût calculatoire plus faible et ses meilleures performances, à condition de bien choisir les longueurs de la permutation et du message. L'association de deux PCC au décodage permet en effet de conserver la corrélation spatiale entre les pixels. Dans le domaine de la DCT, les coefficients hôtes sont moins corrélés et 2D-PCC perd de son intérêt.

La différence entre les résultats théoriques et les résultats pratiques précédents proviennent de la corrélation entre pixels voisins de l'image. Cette propriété est occultée par le modèle AWGN et il est difficile de trouver un modèle théorique adéquat du fait de la non-stationnarité de l'image. Les parcours d'image permettent cependant de passer d'une image bidimensionnelle à une image mono-dimensionnelle tout en conservant les corrélations locales entre pixels. L'utilisation d'une mise en forme répétition permet ensuite de répartir l'influence du bruit sur chaque bit du message, donc d'éviter les cas les plus défavorables, tout en étalant le spectre grâce à la composante aléatoire des PCC. La contribution du bruit de l'image au décodage est ainsi réduite d'un facteur pouvant atteindre 3. Ces résultats rejoignent la problématique de l'entrelacement, très peu étudiée dans le cadre du tatouage d'image (les seuls travaux existant concernent la robustesse dans le domaine de la DCT à un rognage de groupes connexes de pixels). On a montré que la combinaison de 1D-PCC avec le parcours de Peano-Hilbert et une mise en forme répétition permet d'obtenir les mêmes performances que 2D-PCC. Le parcours de Peano-Hilbert sera donc utilisé par la suite dans les algorithmes utilisant les filtres LPTV dans le domaine spatial.

Il serait intéressant d'étudier le lien entre d'une part, la combinaison du parcours de Peano-Hilbert et de l'entrelacement aléatoire PCC et d'autre part, des parcours d'espace aléatoires.

Les filtres LPTV : un outil puissant pour le tatouage par étalement de spectre

Les filtres LPTV sont un ensemble de filtres plus général que les PCC. Ils en partagent cependant les propriétés de périodicité, qui sont utiles dans le domaine spatial avec une mise en forme répétition. Deux types de filtres LPTV inversibles et étalant le spectre ont été introduits. Les LL-LPTV, définis par leurs matrices polyphases, ont déjà été utilisés dans le domaine des télécommunications et sont inspirés de travaux sur les bancs de filtres à décimation maximale. Nous avons également proposé un ensemble de filtres LPTV à filtres modulateurs constants par morceaux, baptisé mod-LPTV. Celui-ci peut être étendu au tatouage multi-utilisateur. Dans leur version de base, ces deux techniques présentent des performances similaires à PCC+Peano, donc supérieures à celles de DS dans le domaine spatial.

Deux autres types de filtres LPTV ont ensuite été définis pour tirer parti des potentialités des filtres LPTV dans l'application au tatouage. ZI-LPTV, qui permet de se prémunir des interférences à bande étroite, est une sorte de "tatouage à spectre de l'hôte connu", par analogie avec le "tatouage à statistique de l'hôte connue" et le "tatouage à état de l'hôte connu" (tatouage informé) qui visent habituellement à annuler les interférences dans le domaine d'insertion. Ici, on impose au filtre LPTV inverse d'annuler les interférences à bande étroite dans la bande du message. Cette idée n'est pas nouvelle puisqu'il est possible de combiner la technique DS avec une mise en forme NRZ et un filtre passe-haut avant décodage, mais son application aux filtres LPTV est élégante et

offre de bonnes performances, notamment face au rognage, malgré la mise en forme NRZ. Les performances expérimentales de ZI-LPTV sont très bonnes, même si la non-stationnarité de l'image rend impossible une pré-annulation complète des interférences dans le domaine spectral. ZI-LPTV divise la contribution d'une image hôte au décodage jusqu'à un facteur 100, ou de manière équivalente, permet de multiplier par 100 la charge utile en l'absence de bruit. La seconde technique impose des contraintes sur le spectre du tatouage simultanément à l'étalement, afin de respecter la "Contrainte du Spectre de Puissance". Cette méthode de tatouage, appelée mask-LPTV, a surtout pour but d'illustrer les potentialités des filtres LPTV par rapport à une modulation simple. Nous avons également montré que les résultats sur le niveau de sécurité théorique des techniques DS sont applicables aux filtres LPTV.

Il est enfin à noter que si les filtres LPTV sont un cadre théorique assez complexe, les algorithmes de tatouage proposés dans cette thèse (LL-LPTV, mod-LPTV, orth-LPTV, ZI-LPTV, mask-LPTV) sont extrêmement simples d'implantation. Des filtres LPTV plus sûrs et présentant des propriétés temporelles particulières pourraient être construits en introduisant des retards dans les filtres mod-LPTV et LL-LPTV.

Combinaison avec les techniques de tatouage classiques

Nous avons montré que les PCC peuvent être considérés comme une mise en forme du message avant application d'un code DS. L'utilisation d'un code DS est ici inutile du point de vue de l'étalement, mais augmente la sécurité de l'algorithme. Cependant, elle élimine les particularités des PCC sur des images naturelles.

D'autre part, nous avons étudié la combinaison des PCC et des filtres LPTV avec d'autres éléments de la chaîne de tatouage. Lorsqu'on utilise les propriétés statistiques de l'image, les techniques proposées perdent une partie de leur spécificité. Cependant, elles offrent les mêmes performances que les techniques de type DS. La combinaison de PCC+Peano ou des filtres LPTV avec l'étalement de spectre amélioré dans le domaine spatial permet d'améliorer la charge utile maximale où le rejet des interférences de l'hôte est total. Enfin, il est possible de réaliser un tatouage quantificatif à transformation d'étalement fondé sur les PCC et les filtres LPTV.

Une perspective à ces travaux serait de proposer des motifs de resynchronisation de type *template* (cf. paragraphe 1.5.4) générés et détectés à l'aide de filtres LPTV.

Bibliographie

[AV00] S. Akkarakaran and P.P. Vaidyanathan. Bifrequency and bispectrum maps : A new look at multirate systems with stochastic inputs. *IEEE Trans. on Signal Processing*, 48(3) :723–736, 2000.

[Bar05] M. Barni. Effectiveness of exhaustive search and template matching against watermark desynchronization. *IEEE Signal Proc. Letters*, 12(2) :158–161, 2005.

[BBV98] M. Blaum, J. Bruck, and A. Vardy. Interleaving schemes for multidimensional cluster errors. *IEEE Transactions on Information Theory*, 44 :730–743, 1998.

[BDBT06] J.-P. Boyer, P. Duhamel, and J. Blanc-Talon. Performance analysis of scalar DC-QIM for watermark detection. *Proc. of ICASSP*, 2006.

[Bou98] J. J. Boutros. Les Turbo Codes Parallèles et Séries. *support de cours ENST*, 1998.

[CB01] C. Coltman and A. Bors. Hierarchical watermarking depending on local constraints. *Proc. of ICIP*, pages 1011–1014, 2001.

[CCL+04] W. Chauvet, B. Cristea, B. Lacaze, D. Roviras, and A. Duverdier. Design of orthogonal LPTV filters : Application to spread spectrum multiple access . *Proc. of ICASSP*, 2004.

[CEL+04] B. Cristea, B. Escrig, B. Lacaze, D. Roviras, and W. Chauvet. Synchronization algorithm for LPTV-based spread spectrum signals . *Proc. of EUSIPCO*, 2004.

[CFF05] F. Cayre, C. Fontaine, and T. Furon. Watermarking security : Theory and practice. *IEEE Trans. on Signal Processing, Special Issue on Content Protection*, 53(10) :3976–3975, 2005.

[Cha04] W. Chauvet. *Etude des filtres LPTV numériques. Application aux communications numériques*. PhD thesis, Institut National Polytechnique de Toulouse, 2004.

[Cit] City University of Hong Kong Corel Image Database. http ://abacus.ee.cityu.edu.hk/ benjiman/corel_1/.

[CLRD05] W. Chauvet, B. Lacaze, D. Roviras, and A. Duverdier. Analysis Banks, Synthesis Banks, LPTV filters : Proposition of an equivalence definition. Application to the design of invertible LPTV filters. *Proc. of EUSIPCO*, 2005.

[CMB02] I.J. Cox, M.L. Miller, and J.A. Bloom. *Digital Watermarking*. Morgan Kaufmann Publisher, Inc., San Francisco, 2002.

[CR03] M. Coulon and D. Roviras. Multi-user detection for random permutation-based multiple access. *IEEE ICASSP'03, Proc.*, 4 :61–64, 2003.

[CR04] M. Coulon and D. Roviras. MMSE Joint Detection for an Asynchronous Spread-Spectrum System Based on Random Permutations. *IEEE ICASSP'04, Proc.*, 2 :17–21, 2004.

[CRE05] B. Cristea, D. Roviras, and B. Escrig. Multipath effect mitigation in LPTV-based multiple access system. *Proc. of EUSIPCO*, 2005.

[Cri04] B. Cristea. *Techniques d'accès multiple avec les changements d'horloge périodiques*. PhD thesis, Institut National Polytechnique de Toulouse, 2004.

[DCOM00] R. Dafner, D. Cohen-Or, and Y. Matias. Context-based space filling curves. *Computer Graphics Forum*, 19(3), 2000.

[dCTG02] L. de Campos Teixeira Gomes. *Tatouage de signaux audio*. PhD thesis, Université René Descartes - Paris V, 2002.

[Elm99] G.F. Elmasry. *Detection and robustness of digital image watermarking signals : a communication theory approach*. PhD thesis, New Jersey Institute of Technology, Newark, NJ, 1999.

[FD03] T. Furon and P. Duhamel. An asymmetric watermarking method. *IEEE. Trans. on Signal Proc.*, 51(4) :981–995, 2003.

[FG99] J. Fridrich and M. Goljan. Protection of digital images using self embedding. *Symp. Content Security and Data Hiding in Digital Media, New Jersey Institute of Technology*, 1999.

[Gar94] W. A. Gardner. *Cyclostationarity in Communications and Signal Processing*. New York : IEEE Press, 1994.

[HPG99] J.R. Hernández and F. Pérez-González. Statistical analysis of watermarking schemes for copyright protection of images. *IEEE Proc., Special Issue on Identification and Protection of Multimedia Information*, 87(7) :1142–1166, 1999.

[HW99] C.-T. Hsu and J.-L. Wu. Hidden Digital Watermarks in Images. *IEEE Transactions on Image Processing*, 8(1) :58–68, 1999.

[KM03] D. Kirovski and H.S. Malvar. Spread-spectrum watermarking of audio signals. *IEEE Trans. on Signal Processing*, 51(4) :1020–1033, 2003.

[Lac96] B. Lacaze. Stationary clock changes on stationary processes. *Signal Processing*, 55 :191–205, 1996.

[Lac00] B. Lacaze. *Processus aléatoires pour communications numériques*. Hermès, 2000.

[LB84] C. Loeffler and C. Burrus. Optimal design of periodically time-varying and multirate digital filters. *Acoustics, Speech, and Signal Processing, IEEE Transactions on*, 32(5) :991 – 997, 1984.

[LR02] B. Lacaze and D. Roviras. Effect of random permutations applied to random sequences and related applications. *Signal Processing*, 82 :821–831, 2002.

[McL99] D. McLernon. One-dimensional Linear Periodically Time-Varying structures : derivations, interrelationships and properties. *IEE Proc.-Vis. Image Signal Proc.*, 146(5) :245–252, 1999.

[PFCPG05] L. Pérez-Freire, P. Comesaña, and F. Pérez-González. Detection in quantization-based watermarking : performance and security issues. *SPIE*, 2005.

[PFCTPPG06] L. Pérez-Freire, P. Comesaña, J.R. Troncoso-Pastoriza, and F. Pérez-González. Watermarking security : a survey. *LNCS Transactions on Data Hiding and Multimedia Security. To appear.*, 2006.

[RL87] T. Rohlev and C. Loeffler. Invertible Periodically Time-Varying Digital Filters. *Proc. of ICASSP*, pages 2380–2382, 1987.

[RLT02] D. Roviras, B. Lacaze, and N. Thomas. Effects of Discrete LPTV on Stationary Signals. *IEEE ICASSP'02, Proc.*, 2 :1127–1220, 2002.

[SEG01] J. Su, J. Eggers, and B. Girod. Analysis of digital watermarks subjected to optimum linear filtering and additive noise. *Signal Processing*, 81 :1141–1175, 2001.

[SG02] J.K. Su and B. Girod. Power-spectrum condition for energy-efficient watermarking. *IEEE Trans. on Multimedia*, 4 :551–560, 2002.

[Tre68] H.L. Van Trees. *Detection, Estimation and Modulation Theory*. New York : Wiley, 1968.

[TSN+01] S. Tsekeridou, V. Solachidis, N. Nikolaidis, A. Nikolaidis, A. Tefas, and I. Pitas. Statistical analysis of a watermarking system based on bernoulli chaotic sequences. *Signal Processing*, 81(6) :1273–1293, 2001.

[Vai87] P.P. Vaidyanathan. Theory and design of M-channel maximally quadrature mirror filters with arbitrary M having the perfect reconstruction

property. *IEEE Trans. on Acoustics, Speech and Signal Processing*, 37 :378–389, 1987.

[Ver98] S. Verdu. *Multiuser Detection*. Cambridge University Press, 1998.

[Vet89] M. Vetterli. Invertibility of linear periodically time-varying filters. *Circuits and Systems, IEEE Transactions on*, 36(1) :148–150, 1989.

[VP98a] G. Voyatzis and I. Pitas. Chaotic watermarks for embedding in the spatial digital image domain. *Proc. of ICIP*, pages 432–436, 1998.

[VP98b] G. Voyatzis and I. Pitas. Digital image watermarking using mixing systems. *Computer and Graphics*, 22(3), 1998.

[Wes05] Andreas Westfeld. Space filling curves in steganalysis. *SPIE Security, Steganography, and Watermarking of Multimedia Contents VII*, 2005.

[Z. 05a] Z. Yücel and A. Bülent Özgüler. An audio watermarking algorithm via zero assigned filter banks. *Proc. of EUSIPCO 05*, 2005.

[Z. 05b] Z. Yücel and A. Bülent Özgüler. An image watermarking algorithm via zero assigned filter banks. *Proc. of IEEE Symposium on Signal Processing and Information Technology*, 2005.

Chapitre 3

Liens entre interpolation et tatouage numérique

Sommaire

Le chapitre précédent concernait principalement les "attaques d'effacement" sur la robustesse, qui ont pour but d'enlever ou de noyer sous un bruit le tatouage afin d'en empêcher la détection. Les attaques géométriques, quant à elles, recouvrent un certain nombre de traitements usuels du document tels que les transformations affines globales ou locales. Sans enlever le tatouage, elles introduisent une désynchronisation entre le tatouage et le décodeur. Pour une distorsion perceptuellement faible et sans modifier la sémantique du document, le décodage est souvent impossible. Nous insistons dans cette partie sur la présence d'un bruit d'interpolation après un enchaînement désynchronisation-resynchronisation. Tout d'abord, nous présentons un état de l'art des techniques d'interpolation. La robustesse des techniques de tatouage fondés sur les filtres LPTV est ensuite étudiée. Enfin, nous proposons de construire un ensemble de masques perceptuels pour les techniques de tatouage classiques, mettant à profit les propriétés perceptuelles de l'interpolation.

3.1 Présentation des techniques d'interpolation

Cette partie présente un état de l'art des techniques d'interpolation. Nous proposons ensuite de modéliser l'erreur d'interpolation d'une image par une distribution gaussienne généralisée (GGD). Ce modèle sera utile dans le chapitre 4. Les techniques d'interpolation utilisées en pratique dans les chapitres 3 et 4 sont essentiellement l'in-

terpolation bilinéaire et par spline bicubique. Cependant, les autres techniques d'interpolation présentées permettent d'envisager des extensions aux techniques de tatouage proposées.

3.1.1 Techniques d'interpolation

On appelle interpolation la construction d'une fonction continue à partir de mesures discrètes en des points donnés. La fonction continue construite doit correspondre aux données discrètes pour les points où les mesures sont disponibles. On suppose également que la représentation continue existe, et la solution de l'interpolation n'est valable que pour un modèle donné du signal. Ainsi, la reconstruction parfaite de fonctions polynomiales, rationnelles ou de polynômes trigonométriques à partir d'un certain nombre de leurs valeurs (et de quelques conditions supplémentaires) est possible et a été étudiée entre autres par Newton, Hermite ou Birkhoff. Cependant, en traitement du signal, les données étudiées suivent rarement des modèles mathématiques aussi simples. On utilise donc d'autres modèles, notamment celui des signaux à bande limitée, pour construire d'autres techniques d'interpolation. Si la fonction à reconstruire ne suit pas le modèle considéré, la reconstruction par interpolation ne sera pas parfaite. L'interpolation peut donc créer des artefacts : oscillations là où l'on attend des valeurs constantes, repliement, effet de bloc (si les fonctions de synthèse définies dans le suite sont à support fini), effet de flou (conséquence du repliement) [TBU00].

En traitement d'images, les données de départ sont souvent équiréparties, notamment lorsqu'on effectue des transformations géométriques (rotations, changements d'échelle). Lorsque les points sont équirépartis, on appellera ici "grille d'interpolation" l'ensemble des données de départ. On peut trouver dans [TBU00] une comparaison rigoureuse de l'ensemble des méthodes d'interpolation classiques en traitement du signal (PPV, linéaire, splines cubiques et B-splines avec divers noyaux...) en fonction de leur déformation visuelle (mesurée par le SNR) et du temps d'exécution. Cette section détaille les techniques d'interpolation au plus proche voisin, linéaire, par convolution quadratique ou cubique, et par splines.

Dans la suite, on notera $\hat{x}(t)$ ou $\hat{x}(t_1, t_2)$ le résultat d'une interpolation à partir de $\mathbf{x}$ au point t ou (t_1, t_2). Si $\mathbf{x}$ résulte de l'échantillonnage d'un signal continu, on notera $x(t)$ ce signal original.

Interpolation au plus proche voisin

Cette méthode consiste à affecter au point interpolé la valeur de son voisin le plus proche. Si les voisins sont équidistants, on doit choisir, par exemple, le plus proche voisin à gauche. Cette méthode est souvent peu satisfaisante perceptuellement.

Interpolation linéaire et bilinéaire

L'interpolation linéaire consiste à affecter au point interpolé la moyenne pondérée des valeurs prises par les 2 points les plus proches. La pondération est inversement proportionnelle à la distance entre le point à interpoler et son voisin :

$$\hat{x}(t) = \frac{t - k_1}{k_2 - k_1} x_{k_1} + \frac{k_2 - t}{k_2 - k_1} x_{k_2}$$

L'extension à 2 dimensions, ou interpolation bilinéaire, consiste à effectuer la moyenne pondérée sur les 4 points les plus proches. Si les points de la grille sont

équirépartis, cela consiste à appliquer successivement une interpolation linéaire sur les lignes puis sur les colonnes (cf. *fig.* 3.1). Il s'agit d'un produit tensoriel. L'expression générale est donc la suivante (cas équiréparti).
Interpolation linéaire sur les lignes :

$$\hat{x}(t_1, k_{1,2}) = \frac{t_1 - k_{1,1}}{k_{1,2} - k_{1,1}} x(k_{1,1}, k_{1,2}) + \frac{k_{2,1} - t_1}{k_{2,1} - k_{1,1}} x(k_{2,1}, k_{1,2})$$
$$\hat{x}(t_1, k_{2,2}) = \frac{t_1 - k_{1,1}}{k_{2,1} - k_{1,1}} x(k_{1,1}, k_{2,2}) + \frac{k_{2,1} - t_1}{k_{2,1} - k_{1,1}} x(k_{2,1}, k_{2,2})$$

puis interpolation linéaire sur les colonnes :

$$\hat{x}(t_1, t_2) = \frac{t_2 - k_{1,2}}{k_{2,2} - k_{1,2}} \left(\frac{t_1 - k_{1,1}}{k_{2,1} - k_{1,1}} x(k_{2,1}, k_{2,2}) + \frac{k_{2,1} - t_1}{k_{2,1} - k_{1,1}} x(k_{1,1}, k_{2,2}) \right) + \frac{k_{2,2} - t_2}{k_{2,2} - k_{1,2}} \left(\frac{t_1 - k_{1,1}}{k_{2,1} - k_{1,1}} x(k_{2,1}, k_{1,2}) + \frac{k_{2,1} - t_1}{k_{2,1} - k_{1,1}} x(k_{1,1}, k_{1,2}) \right)$$

L'interpolation bilinéaire, même si elle est très simple, donne souvent des résultats très satisfaisants visuellement.

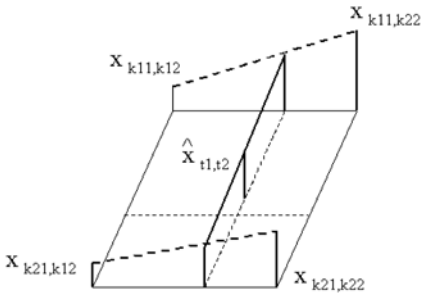

FIG. 3.1 – Interpolation bilinéaire

Interpolation par splines

Une spline d'ordre n est une fonction polynomiale par morceaux dont chaque polynôme est de degré n, et qui présente des propriétés de régularité (continuité de ses dérivées). L'interpolation par spline consiste à construire une fonction $\hat{x}(t)$ comme le résultat de la superposition de fonctions de synthèse f^n centrées sur les points de la grille et pondérées. Le support de f^n englobe les $(n+1)$ points les plus proches. Dans la suite, on se ramènera pour plus de simplicité à la grille $\{k \in \mathbb{Z}\}$, donc à des points de départ équirépartis $\{x_k\}_{k\in\mathbb{Z}}$. Le support de f^n est donc $[-\frac{n+1}{2}, \frac{n+1}{2}]$ [Uns99].
L'interpolation est dite du 1er type ou "interpolation par convolution" si :

$$\hat{x}(t) = \sum_{k=-\infty}^{+\infty} x_k f^n(t-k)$$

et du 2nd type ou "interpolation généralisée" si :

$$\hat{x}(t) = \sum_{k=-\infty}^{+\infty} c_k f^n(t-k)$$

Dans ce dernier cas, les c_k sont des coefficients calculés à partir des x_k, ce qui implique un coût calculatoire supérieur à celui de l'interpolation de type 1. L'emploi du terme "spline" pour désigner à la fois l'interpolation par convolution et l'interpolation généralisée prête parfois à confusion. Par exemple, les B-splines définies dans le suite ne sont pas performantes si elles sont utilisées dans une interpolation du premier type.

Les contraintes d'interpolation varient. Elles peuvent inclure la "condition d'interpolation" : $\forall k \in \mathbb{Z}, x(k) = x_k$, ainsi que des contraintes de régularité (*i.e.* de continuité des dérivées jusqu'à un ordre donné). Ceci se ramène à un système d'équations dérivé des conditions de dérivabilité aux nœuds.

La plupart des techniques d'interpolation sont des fonctions linéaires de $\{x_k\}_{k\in\mathbb{Z}}$. La technique dite d'"interpolation linéaire" revient de plus à utiliser la fonction $f^n(t) = 1 - |t|, t \in [-1, 1], f^n(t) = 0, t \notin [-1, 1]$ qui est linéaire en t, d'où son nom.

Interpolation par spline convolutive

L'interpolation par spline convolutive d'ordre $n = 0$ équivaut à une interpolation par la méthode des plus proches voisins :

$$\hat{x}(k+h) = \begin{cases} x_k \text{ si } 0 < h < 1/2 \\ x_{k+1} \text{ si } 1/2 < h < 1 \end{cases}$$

L'interpolation par spline convolutive d'ordre 1 revient à une interpolation linéaire :

$$\hat{x}(k+h) = (1-h)x_k + hx_{k+1}$$

Malgré son efficacité, la spline convolutive d'ordre 2 (dont le support englobe les 3 points les plus proches) est moins utilisée pour des raisons de dissymétrie [Dod97].
L'ordre 3 est le plus utilisé. La spline convolutive cubique est une spline d'ordre 3 dont la fonction de synthèse est composée de deux polynômes distincts. C'est donc une spline polynomiale par morceaux. Les polynômes sont choisis d'après les conditions d'interpolation, ce qui conduit selon les auteurs à différentes fonctions. Les plus utilisées sont celles de Catmull-Rom [CR74] et Keys [Key81]. Par exemple, la spline cubique de Keys répond aux contraintes suivantes :

- condition d'interpolation
- $\hat{x}'(t)$ continue, $\hat{x}''(t)$ pas nécessairement continue
- ordre d'approximation maximal pour le type 1 : si Δ est la distance entre deux nœuds, $\|\hat{x} - x\| = O(\Delta^3)$ *i.e.* les 3 premiers termes du développement de Taylor coïncident.

Interpolation par B-splines

Les Basic-splines ou B-splines [Uns99], notées β^n, sont des polynômes d'ordre n continûment différentiables jusqu'à l'ordre $n-1$. Les B-Splines font partie des splines polynomiales, par opposition aux splines polynomiales par morceaux. Elles ont les propriétés suivantes :

- toute spline polynomiale est une combinaison linéaire de B-splines
- les B-splines sont les splines polynomiales de support le plus court (pour un ordre donné n) et de plus grande régularité
- les B-spline offrent l'ordre d'approximation le plus élevé ($n+1$) pour un support donné de taille ($n+1$).

- les B-splines sont construites par convolution (cf. *fig.*3.2) :

$$\beta^n(t) = (\overbrace{\beta^0 * \cdots * \beta^0}^{(n+1)\text{fois}})(t)$$

- les B-splines réalisent une interpolation de type 2.
- β^1 correspond à une interpolation par les plus proches voisins et β^2 à une interpolation linéaire.

La difficulté d'implantation réside dans le calcul des c_k. Si les points sont équirépartis de pas d'échantillonnage $1/m$, on peut se ramener à un filtrage des points $x_k = \hat{x}(k)$ connus par $\frac{1}{B_m^n(z)}$, où $b_m^n(k) = \beta^n(k/m)$ (B-spline discrète) et $B_m^n(z) = \text{TZ}(b_m^n(k))$. En effet $\hat{x}(k) = (c * b_m^n)_k$, donc $G = CB_m^n$. Il existe également une implantation efficace de ce filtre.

L'interpolation 1D peut être étendue au cas 2D par produit tensoriel pour une application aux images [Uns99] :

$$x(t_1, t_2) = \sum_k \sum_l c_{k,l} \beta^n(x-k) \beta^n(y-l)$$

Les $c_{k,l}$ sont calculés à partir des $x_{k,l}$ en effectuant une interpolation 1D sur les lignes suivie d'une interpolation 1D sur les colonnes.

Les splines d'approximation sont utilisées en débruitage par exemple. Au lieu de respecter la contrainte d'interpolation, elles minimisent l'erreur d'interpolation sous la contrainte de régularité [Uns99] :

$$\sum_{k \in \mathbb{Z}} (x_k - \hat{x}(k))^2 + \lambda \int_{-\infty}^{\infty} (\hat{x}^{(\frac{n+1}{2})}(t))^2 \mathrm{d}t$$

La régularité est ici privilégiée au détriment du respect de la condition d'interpolation.

La grille équirépartie offre des facilités d'implantation de l'interpolation par B-splines. En 1D, lorsque les points de la grille d'interpolation ne sont pas équirépartis, il est possible d'obtenir un équivalent de tous les résultats précédents, mais les calculs sont plus complexes. On calcule en effet les c_k en résolvant un système d'équations par divers algorithmes de factorisation (cf. les travaux de Cox et DeBoor notamment). Les fonctions splines de base, équivalents des B-splines incluant dans leur support les $(n+1)$ points connus les plus proches, existent aussi dans ce cas, mais dépendent du point considéré sur la grille. La spline d'ordre n est calculée en fonction des splines d'ordre $n-1$, mais ici la convolution est remplacée par une relation de récurrence. L'extension à la 2D est plus délicate. L'implantation par produit tensoriel n'est possible que si chaque ligne, chaque colonne a le même espacement (en tissu écossais ou "tartan", cf. *fig.* 3.3). L'interpolation à partir de points dispersés est donc difficile [LWS97].

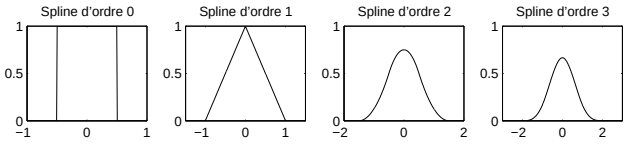

FIG. 3.2 – B-splines d'ordre 0 à 3

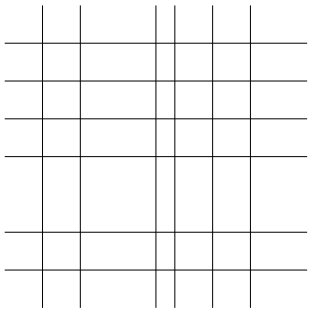

FIG. 3.3 – Exemple de grille non équirépartie

Interpolation par sinus cardinal et B-spline cardinale

Shannon a montré qu'un signal à bande limitée, échantillonné à une fréquence suffisamment grande, peut être reconstruit de façon parfaite par convolution avec un sinus cardinal. Cette reconstruction correspond à un problème d'interpolation. Ce filtrage revient dans le domaine fréquentiel à éliminer les répétitions du spectre dues à l'échantillonnage. Cependant, la fonction sinC décroît lentement, ce qui implique en pratique un coût calculatoire trop élevé. La fonction sinC est donc tronquée par des fenêtres (Dirichlet, Hanning...) et l'interpolation n'est plus parfaite. De plus, les signaux à bande limitée doivent être infinis, ce qui n'est pas possible en pratique. Il existe une autre implantation pratique de l'interpolation par sinC, utilisant des glissements de la transformée de Fourier (*shifted DFT*) [Yar97]. Cette technique est particulièrement intéressante car elle est inversible (c'est une isométrie) et offre une très bonne qualité perceptuelle. Les B-splines correspondent elles aussi à une alternative à l'interpolation convolutive par sinC [Uns99]. En effet, l'interpolation par B-spline peut se ramener à une interpolation du premier type par la spline cardinale d'ordre n définie par (cf. *fig.* 3.4) :

$$\eta^n(t) = \sum_k (b_1^n)_k^{-1} \beta^n(t-k)$$

Ici, $(b_1^n)^{-1}(k) = \mathrm{TZ}^{-1}(1/\mathrm{TZ}(b_1^n(k)))$. η^n oscille autour de 0, est à support infini si $n > 2$ et décroît exponentiellement. Lorsque $n \to +\infty$, η^n tend vers un sinus cardinal. Tout en étant d'implantation pratique et efficace, car à support limité, les B-splines permettent donc de réaliser une interpolation par un interpolant de support infini. Celui-ci présente des propriétés théoriques intéressantes et est proche de l'interpolant idéal pour les signaux à bande limitée. Si un signal ne suit pas le modèle à bande limitée, l'interpolation par sinC ou par B-spline entraîne des oscillations dues au repliement, ce qui peut être corrigé par un préfiltrage du signal avant échantillonnage.

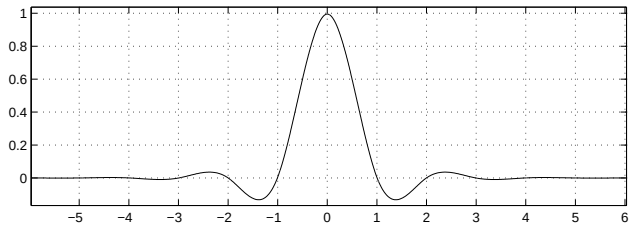

FIG. 3.4 – B-spline cardinale d'ordre 3

Interpolation adaptative

Les techniques d'interpolation adaptative ont pour but de préserver les objets d'une image, et en particulier leurs contours. Elles contrôlent souvent l'interpolation par une détection de contours ou une segmentation de l'image [KS95][GEVS01]. Certaines méthodes itératives utilisent une optimisation par projection sur des ensembles convexes pour la contrainte de continuité des contours. Les techniques d'interpolation orientées contours utilisent un modèle de la source de l'image [LO01]. Enfin, d'autres techniques sont issues de la quantification vectorielle et du filtrage morphologique.

Autres types d'interpolation d'image

Dans le cas d'une interpolation 2D, les techniques évoquées jusqu'ici excluent l'interpolation à partir de données disséminées. Des interpolants à symétrie radiale peuvent alors être utilisés :

$$\hat{x}(t_1, t_2) = \sum_{(k_{1,n}, k_{2,n}) \in \mathcal{G}} x(k_{1,n}, k_{2,n}) \phi(\|(t_1, t_2) - (k_{1,n}, k_{2,n})\|)$$

La populaire "spline de plaque mince" (*thin-plate spline*) correspond à $\phi(r) = r^2 \log(r)$ [Boo89]. Ces techniques sont cependant très gourmandes en temps de calcul, et sont souvent destinées à l'interpolation de surface 3D.

Les techniques d'interpolation par voisins naturels, particulièrement adaptées à des données multidimensionnelles, utilisent un partitionnement de Voronoï (ou triangulation de Delaunay) des données [Sib81]. Deux nœuds partageant une face de leurs cellules de Voronoï respectives sont appelés voisins naturels. Divers interpolants sont utilisés, prenant en compte la valeur des nœuds et les aires des cellules. Dans [PLK$^+$06], on trouve un exemple d'interpolation d'une image 2D à partir de points disséminés choisis aléatoirement ou de façon adaptée à l'image.

Application à l'image : exemples visuels

Les *fig.* 3.6, 3.7 et 3.8 montrent des exemples d'interpolation au plus proche voisin, bilinéaire et par spline à partir de pixels situés sur une grille en quinconce.

FIG. 3.5 – Détail de Lena

FIG. 3.6 – Exemple d'interpolation au plus proche voisin (Lena)

FIG. 3.7 – Exemple d'interpolation bilinéaire (Lena)

FIG. 3.8 – Exemple d'interpolation par splines (Lena)

3.1.2 Etude de l'erreur d'interpolation d'image : modèle GGD

Afin d'évaluer la performance d'une technique d'interpolation, il est intéressant d'évaluer l'erreur d'interpolation. Cette étude sera également nécessaire dans le chapitre 4. L'erreur d'interpolation en image est très difficile à prévoir, principalement en raison du manque de modèle théorique convenant aux images naturelles. Les études théoriques de l'erreur d'interpolation concernent des cas particuliers. Dans le cas de l'interpolation bilinéaire, l'erreur peut par exemple être majorée en fonction des dérivées du signal original. Dans le cas des splines, les performances de l'interpolation peuvent être calculées de manière théorique pour un modèle de Markov du premier ordre "représentatif d'une grande variété d'images réelles" ([TBU00] p.27) par intégration du spectre de la fonction interpolante [TBU00]. Afin d'avoir une meilleure adéquation avec les performances pratiques, on propose donc dans la suite d'étudier empiriquement l'erreur d'interpolation pour une méthode, une grille et une image données. La grille d'interpolation utilisée dans les simulations qui suivent est la grille en "quinconce" (cf. partie 4.4.1), et les coordonnées des points interpolés sont déterminées aléatoirement à la manière de W-bilin (cf. partie 4.4.2). Nous noterons l'erreur d'interpolation $\epsilon(\mathbf{x})$.

Modèle de l'erreur d'interpolation

On utilise le modèle gaussien généralisé (GGD), déjà présenté dans la partie 1.5.2, pour modéliser l'erreur d'interpolation $\epsilon(x_k)$ en tout point $k \in \{1, \ldots, N\} \setminus \mathcal{G}$:

$$f_x(x) = Ae^{-|\beta x|^c} \quad , x \in \mathbb{R}$$

avec A et β calculés en fonction de l'écart type σ et de la courbure c :

$$\beta = \frac{1}{\sigma}\left(\frac{\Gamma(3/c)}{\Gamma(1/c)}\right)^{1/2} \qquad A = \frac{\beta c}{2\Gamma(1/c)} \quad .$$

Etude de l'histogramme pour l'interpolation bilinéaire : la *fig.* 3.9 montre que le modèle gaussien correspond assez mal à l'histogramme de l'erreur d'interpolation. Sur les *fig.* 3.10 et 3.11, on constate par contre que le modèle GGD convient bien. c est calculé par un estimateur du maximum de vraisemblance [HPG99]. Expérimentalement, on observe des valeurs de c proches de 0.85 ou 0.9 ($c = 2$ pour une gaussienne, $c = 1$ pour une laplacienne).

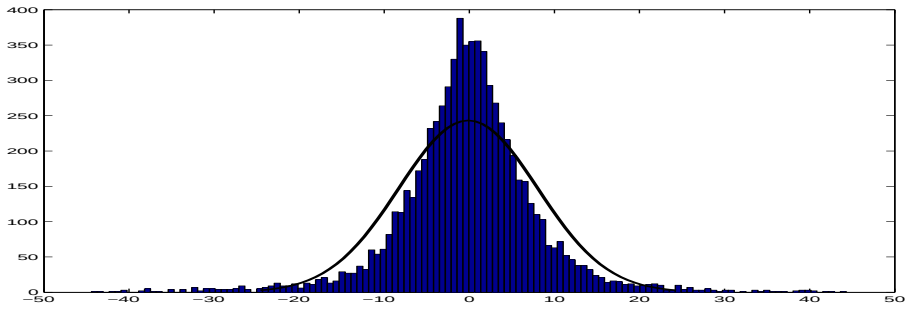

FIG. 3.9 – Histogramme et modèle gaussien de $\epsilon(\mathbf{x})$, bilinéaire, Bateaux

Etude de l'histogramme pour l'interpolation par B-splines bicubiques : les histogrammes sont très proches de ceux obtenus pour l'interpolation bilinéaire (forme

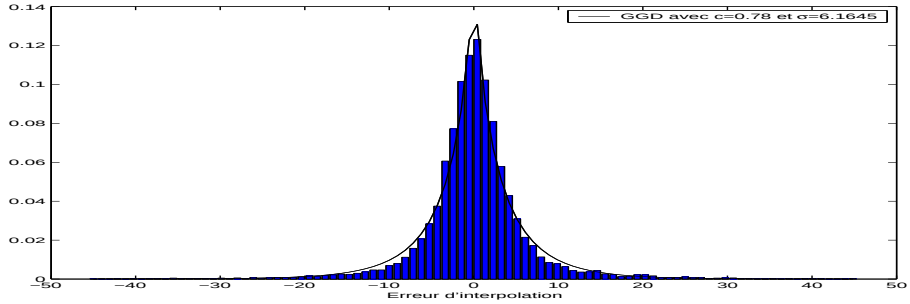

FIG. 3.10 – Histogramme et GGD de $\epsilon(\mathbf{x})$, bilinéaire, Lena

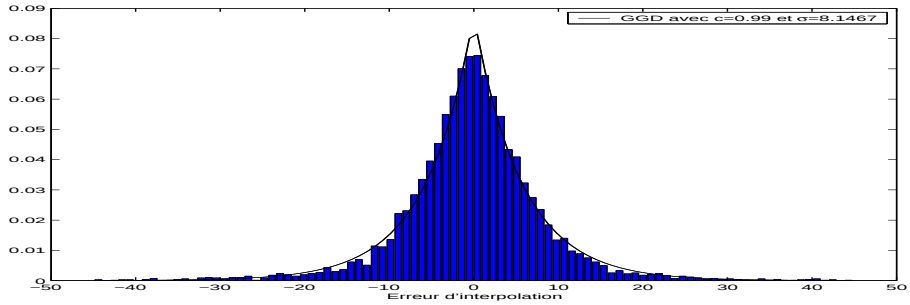

FIG. 3.11 – Histogramme et GGD de $\epsilon(\mathbf{x})$, bilinéaire, Bateaux

de gaussienne généralisée, de paramètre de courbure pour Babouin, Lena, Bateaux : $c = 1.17, 0.81, 1.02$ contre $c = 1.13, 0.78, 0.99$).

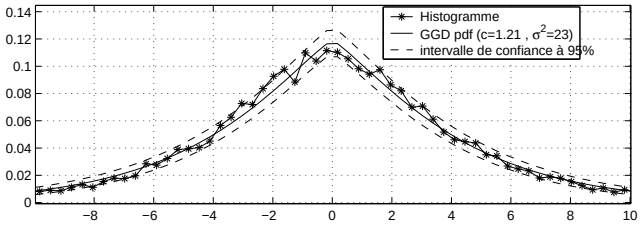

FIG. 3.12 – Histogramme, Lena, spline, 15000 pts

Influence de la troncature des queues : en enlevant les queues de l'histogramme, le modèle GGD est beaucoup mieux respecté et se situe dans un intervalle de confiance à 95% (cf. Figs. 3.13 et 3.14). L'intervalle de confiance est calculé en ne prenant que 15000 points (correspondant approximativement à DWR=28 dB). La courbure est alors 1.15 (Lena), 1.51 (Babouin), 1.4 (Bateaux), 0.75 (Poivrons), 0.85 (Pentagone). Plus l'on tronque les queues, plus la variance augmente (prise en compte des valeurs extrêmes) et plus c diminue (pente de plus en plus grande).

Cohérence du modèle

Nous n'avons pas trouvé dans la littérature d'exemple explicite d'étude de la distribution de l'erreur d'interpolation. Ce modèle est cependant cohérent avec les travaux d'autres auteurs. Huang et Mumford [HM99], Lee *et al.* [LMH01] ont observé empiriquement que la différence d'intensité entre deux pixels adjacents d'une image naturelle suit une laplacienne généralisée (autre nom de la distribution gaussienne généralisée). Green [Gre02] montre, empiriquement encore, que les images naturelles sont différen-

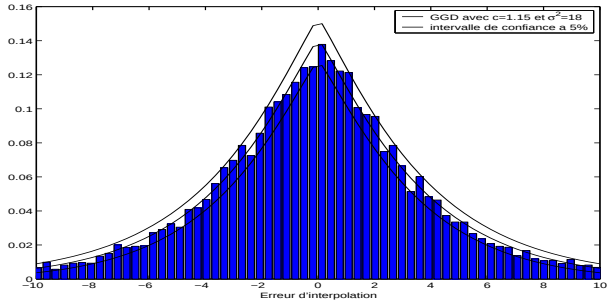

FIG. 3.13 – Histogramme, GGD et intervalle de confiance de $\epsilon(\mathbf{x})$, bilinéaire, Lena, tronqué

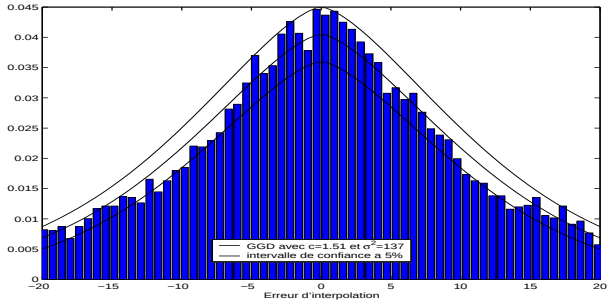

FIG. 3.14 – Histogramme et GGD de $\epsilon(\mathbf{x})$, bilinéaire, Babouin, tronqué

tiellement laplaciennes, *i.e.* une combinaison linéaire de plusieurs pixels adjacents tend à suivre une distribution laplacienne (donc avec une courbure $c = 1$) si la somme des coefficients est nulle. L'interpolant bilinéaire vérifie cette hypothèse. La B-spline cardinale cubique est de support infini, mais décroît exponentiellement. Si on l'approche par un interpolant de support fini (par exemple 4x4), elle vérifie également l'hypothèse.

3.2 Robustesse des filtres LPTV aux attaques désynchronisantes

Toute transformation géométrique est suivie d'un rééchantillonnage destiné à stocker le document sous forme numérique. Aux coordonnées conservées, la valeur des échantillons a préalablement été calculée par interpolation à partir des échantillons du document initial, afin que la déformation perceptuelle soit faible. Le processus d'interpolation-rééchantillonnage s'accompagne forcément d'une perte d'information, même si elle est limitée par les propriétés perceptuelles de l'interpolation. Les effets d'une attaque géométrique ne peuvent donc pas être annulés totalement par la resynchronisation, ce qui est rarement souligné. L'attaque introduit toujours un "bruit d'interpolation", auquel contribue parfois l'opération de resynchronisation elle-même. L'outil de simulation Checkmark propose d'ailleurs une attaque fondée sur un sous-échantillonnage suivi d'une interpolation [Che]. On pourrait également imaginer un modèle d'attaque fondée exclusivement sur l'interpolation. Les attaques désynchronisantes ont été présentées dans le paragraphe 1.2.3.

Par exemple, l'attaque de rotation d'une image est illustrée sur la *fig.* 3.15 [SKH01], [SKH02]. L'attaque crée deux sources d'erreurs. Une forte rotation ($\theta > 1^\circ$) entraîne un changement de coordonnées discrètes. La perte de correspondance a un effet dévastateur sur les performances : il s'agit d'une attaque désynchronisante. Cependant, si θ est grand, l'attaque est perceptible visuellement ainsi que selon le critère du PSNR.

L'angle minimum $\theta_{\min}(k_1, k_2)$ tel qu'un point localisé en $_{k_1,k_2}$ change de position après rotation centrée en $((N_1+1)/2, (N_2+1)/2)$) peut être calculé : si (t_1, t_2) est l'intersection entre le cercle

$$(u-(N_1+1)/2)^2+(v-(N_2+1)/2)^2 = (n_1-(N_1+1)/2)^2+(n_2-(N_2+1)/2)^2 \quad (3.1)$$

et les régions de décision

$$u = n_1 - \frac{1}{2}, u = n_1 + \frac{1}{2}, v = n_1 - \frac{1}{2}, v = n_1 + \frac{1}{2} \ , \quad (3.2)$$

alors $\theta_{\min}(k_1, k_2)$ est l'angle entre les vecteurs $_{k_1,k_2}$ et (t_1, t_2). La proportion des points qui changent de position pour une rotation centrée dépend de l'angle. D'après la *fig.***??**, l'angle minimal pour qu'au moins un point change de position est $\theta = 0.1^\circ$.

Dans le cas d'une rotation de fort angle, DNR=14 dB en moyenne après resynchronisation (pour une rotation utilisant l'interpolation bilinéaire ; on a même DNR=13.8 dB avec l'interpolation bicubique, DNR=11.3 dB avec les plus proches voisins). Donc tout algorithme resynchronisé doit pouvoir supporter un tel bruit. Ce n'est pas toujours le cas de DS (avec $L = 300$, DWR=28 dB, TEB=0.4 pour la rotation et TEB=0.14 pour l'attaque AWGN équivalente). Par contre, DS+W offre de bons résultats quel que soit θ.

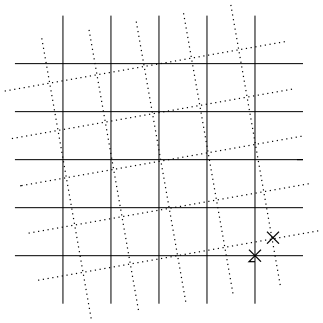

FIG. 3.15 – Exemple de rééchantillonnage par interpolation : rotation d'une image

Les techniques de transformation invariante et d'insertion de mire (cf. paragraphe 1.5.4) sont indépendantes de l'étape Sp d'étalement. Elles visent uniquement à annuler ou inverser les déformations, et ce avant décodage. On peut donc les appliquer indifféremment aux techniques d'étalement de spectre de type DS, PCC ou filtres LPTV. Les PCC et filtres LPTV ne fournissent pas de robustesse supplémentaire aux attaques géométriques, mais les techniques de resynchronisation développées pour le DS restent applicables. L'existence d'un bruit d'interpolation après toute attaque géométrique est particulièrement évidente après implantation d'une technique de resynchronisation telle que celle de [ARPG02] ou de l'insertion dans un domaine spatial invariant tel que la normalisation d'image [DBG$^+$05]. Pourtant, dans les deux articles cités, l'erreur au décodage n'est étudiée que sous l'angle de la désynchronisation résiduelle, et le bruit d'interpolation n'est pas quantifié. Dans l'annexe A.3, nous étudions la robustesse des techniques de tatouage fondés sur les filtres LPTV au bruit d'interpolation résiduel après resynchronisation par [ARPG02]. Le paragraphe 4.6 contient des comparaisons de robustesse de techniques de tatouage informé au bruit d'interpolation.

3.3 Une classe de masques perceptuels utilisant l'interpolation

L'interpolation a jusqu'ici été considérée essentiellement comme une source de nuisance pour le tatouage puisque, que ce soit dans la construction d'un domaine transformé ou lors d'une attaque, elle gêne le décodage du tatouage. Dans ce paragraphe, nous proposons au contraire de construire un ensemble de masques perceptuels mettant à profit les propriétés perceptuelles de l'interpolation.

3.3.1 Rôle de l'interpolation dans un schéma de tatouage

Jusqu'à présent, l'interpolation a été envisagée en tatouage principalement au niveau des attaques. L'interpolation peut également intervenir lors d'une transformée lorsqu'à partir de données discrètes on souhaite se placer dans un domaine d'insertion continu tel que la transformée de Fourier-Mellin [JP98] ou une transformation géométrique [BCM00]. Là encore, il s'agit d'une source d'erreur pour l'estimation du message. En effet, le document tatoué étant nécessairement numérique et à valeurs discrètes, toute interpolation est suivie d'un rééchantillonnage, donc d'une perte d'information. Dans deux cas très particuliers, l'interpolation est utilisée comme élément à part entière de la technique de tatouage.

Une méthode de tatouage (ou même de stéganographie) proposée par Boato *et al* [BFM05] utilise l'interpolation, mais en tant qu'outil de cryptographie et non de traitement du signal. La méthode est hiérarchique (le secret est partagé entre plusieurs utilisateurs qui en réfèrent à une autorité légale) et déterministe (le message secret est reconstruit parfaitement et non estimé). Le message est associé à des entiers par une table de hachage, puis ceux-ci sont associés aux coefficients d'un polynôme trigonométrique. On cache dans l'image les valeurs prises par le polynôme et ses dérivées (correspondant aux niveaux hiérarchiques successifs) en des points donnés par un algorithme additif dans le domaine de la DCT. Au décodage, la fonction polynomiale secrète est reconstruite par interpolation polynomiale de Birkhoff (qui revient à résoudre un système linéaire), ce qui n'est possible qu'à partir d'un nombre suffisamment grand de valeurs du polynôme et de ses dérivées. La connaissance de la fonction continue complète permet de retrouver les coefficients du polynôme, et donc le message.

Dans le contexte du tatouage d'objets 3D, on ne tatoue pas une image ou sa représentation visuelle 3D, mais le modèle mathématique (maillages, surfaces, paramètres) qui sert à générer l'objet graphique par interpolation. De nombreux modèles 3D (surfaces, objets 3D dans les vidéos…) sont représentés par des splines (notamment des NURBS : B-Splines Rationnelles Non-Uniformes). Le tatouage peut agir notamment sur le degré de la spline ou sur l'emplacement des nœuds [OMA99]. On peut trouver un état de l'art du tatouage d'objets 3D dans [DP04]. On tatoue donc la technique d'interpolation en elle-même, ce qui n'est pas applicable au tatouage d'images naturelles.

3.3.2 Proposition de masques perceptuels fondés sur l'interpolation

Dans cette section, on construit des masques perceptuels (cf. Section 1.5.3) pour les techniques de tatouage par étalement de spectre DS et PCC. D'une part, les techniques d'interpolation découlent d'études théoriques sur les signaux à bande limitée, destinées à garantir la qualité perceptuelle. D'autre part, l'erreur d'interpolation permet de faire

ressortir les hautes fréquences d'un signal, ce qui correspond à l'approche classique de la génération d'un masque spatial. Nous proposons donc de **pondérer le tatouage par l'erreur d'interpolation en un point donné**.

Cette proposition s'appuie sur l'hypothèse selon laquelle l'ajout d'une version pondérée de l'erreur d'interpolation sera peu perceptible. En effet, la technique d'interpolation est construite afin de minimiser l'impact de l'erreur interpolation exacte. Nous n'avons donc pas de certitude *a priori* sur le comportement d'une version atténuée de cette erreur, et surtout d'un changement de son signe. Nous vérifions par la suite la validité de cette hypothèse par des expérimentations.

Le principe d'utiliser la différence $g(\mathbf{x}) - \mathbf{x}$ pour pondérer une séquence DS, où $g(\mathbf{x})$ est perceptuellement proche de $\mathbf{x}$, rejoint les travaux de [HJMM03]. Dans [HJMM03], g est une compression avec perte de $\mathbf{x}$ (telle que JPEG2000 avec différents paramètres). Cependant, dans [HJMM03], $g(\mathbf{x})$ est recalculé à partir de $\mathbf{x}$ à la réception : le schéma est non aveugle, ce qui lui enlève une grande part de son intérêt.

Interpolation linéaire

Pour une application à l'image, le calcul de l'erreur d'interpolation bilinéaire se ramène à un filtrage de $\mathbf{x}$ par

$$h_\psi(k,l) = \frac{1}{4}\begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix}$$

ce qui définit un filtre Laplacien qui calcule les dérivées secondes horizontale et verticale de l'image [CP95]. Cette interprétation est particulièrement intéressante puisqu'un masque spatial classique utilise également les dérivées secondes horizontale, verticale mais aussi diagonales [KJ99] (cf. paragraphe 1.5.3). C'est également une extension aux dérivées secondes du masque de [ARPG02]. Le spectre du masque est proche de celui de l'image originale, ce qui est intéressant pour l'imperceptibilité et la robustesse au débruitage [VHBP99]. Cependant, ce masque est moins performant que le masque laplacien de [KJ99] puisqu'il omet les diagonales.

Interpolation par spline cubique

Pour le cas d'une interpolation par spline bicubique, le calcul de l'erreur d'interpolation en un point donné est plus délicat (cf. partie 3.1). En effet, du fait de la condition d'interpolation, le point (k_1, k_2) ne doit pas appartenir à la grille d'interpolation. Donc celle-ci ne peut pas à la fois être répartie de façon égale sur les lignes et les colonnes (*i.e.* non dispersée, cf. Section 3.1) et inclure tous les points voisins connus. Une première solution consiste à utiliser un voisinage en échiquier comme grille d'interpolation pour les B-splines cubiques (cf. partie 4.4.3) (**masque bspline**). Le filtre h_ψ peut alors être approché par :

$$\begin{bmatrix} 0 & 0.0014 & 0 & 0.0011 & 0 & 0.001 & 0 & 0.0011 & 0 & 0.0014 & 0 \\ 0.0014 & 0 & -0.0051 & 0 & -0.004 & 0 & -0.004 & 0 & -0.0051 & 0 & 0.0014 \\ 0 & -0.0051 & 0 & 0.0195 & 0 & 0.0152 & 0 & 0.0195 & 0 & -0.0051 & 0 \\ 0.001 & 0 & 0.0195 & 0 & -0.0735 & 0 & -0.0735 & 0 & 0.0195 & 0 & 0.001 \\ 0 & -0.004 & 0 & -0.0735 & 0 & 0.3568 & 0 & -0.0735 & 0 & -0.004 & 0 \\ 0.001 & 0 & 0.0152 & 0 & 0.3568 & -1 & 0.3568 & 0 & 0.0152 & 0 & 0.001 \\ 0 & -0.004 & 0 & -0.0735 & 0 & 0.3568 & 0 & -0.0735 & 0 & -0.004 & 0 \\ 0.001 & 0 & 0.0195 & 0 & -0.0735 & 0 & -0.0735 & 0 & 0.0195 & 0 & 0.001 \\ 0 & -0.0051 & 0 & 0.0195 & 0 & 0.0152 & 0 & 0.0195 & 0 & -0.0051 & 0 \\ 0.0014 & 0 & -0.0051 & 0 & -0.004 & 0 & -0.004 & 0 & -0.0051 & 0 & 0.0014 \\ 0 & 0.0014 & 0 & 0.0011 & 0 & 0.001 & 0 & 0.0011 & 0 & 0.0014 & 0 \end{bmatrix}$$

FIG. 3.16 – Masque utilisant l'interpolation bilinéaire, Lena

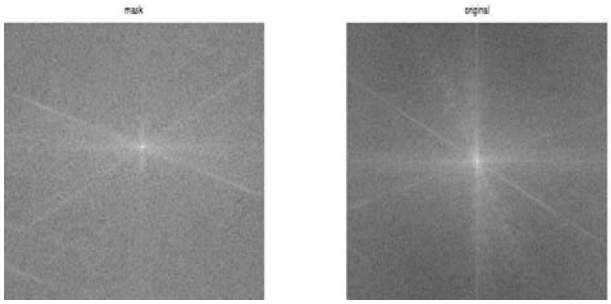

FIG. 3.17 – Bateaux : spectre du masque bilinéaire (g) et de l'image originale (d)

L'interpolation bénéficie des performances des B-splines cubiques. Cependant, les performances pâtissent du support en échiquier : certains points du voisinage sont omis, alors que tout le voisinage est connu. De la même manière, on peut utiliser des splines cubiques convolutives sur un voisinage en échiquier [PF05] (**masque cspline**) :

$$h_\Psi = \frac{1}{256}\begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -9 & 0 & -9 & 0 & 0 \\ 0 & -9 & 0 & 81 & 0 & -9 & 0 \\ 1 & 0 & 81 & -256 & 81 & 0 & 1 \\ 0 & -9 & 0 & 81 & 0 & -9 & 0 \\ 0 & 0 & -9 & 0 & -9 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

Une autre implantation consiste à exploiter tous les points du document pour calculer les coefficients passe-haut du signal, en effectuant un sous-échantillonnage (changement d'échelle de facteur $\delta = 0.5$) suivi d'un suréchantillonnage (changement d'échelle de facteur $\delta = 2$), tous deux par interpolation par B-spline cubique. Nous appellerons cette technique **masque scaling**.

FIG. 3.18 – Masque bspline, Lena

FIG. 3.19 – Masque cspline, Lena

FIG. 3.20 – Masque scaling, Lena

Application à l'image : comparaisons perceptuelles

Les distances de Watson, SSIM, Kullback-Leibler, CG, C4 et Komparator sont utilisées pour évaluer les différents masques à DWR (donc EQM et PSNR) donnés sur la base de 44 images [Cit]. Il suffit d'étaler par une séquence $\Psi\mathbf{c}$ pour effectuer un masque perceptuel dans une technique quantificative avec redondance par étalement.

Avec la redondance par répétition, on doit recourir à la quantification adaptative. Le tableau ci-dessous présente les résultats expérimentaux des masques proposés, comparés aux masques traditionnels.

Cette étude confirme les très bonnes performances perceptuelles des trois masques proposés. Comme prévu, le masque laplacien est meilleur que le masque d'interpolation bilinéaire et celui inspiré de W-spline. Le masque scaling est meilleur que ceux de Laplace et Alvarez selon les critères SSIM, Kullback-Leibler, CG et Watson. Il s'agit donc d'un bon choix de masque perceptuel. De plus, des simulations sur la robustesse suggèrent que celle-ci bénéficie de l'emploi d'un masque. En effet, le calcul des hautes fréquences à la réception effectue un débruitage local (similaire au préfiltrage de Wiener) et le corrélation entre le tatouage et l'image améliore la robustesse au débruitage. Ces propriétés sont déjà constatées pour les masques laplaciens et NVF.

Les masques NVF et DCT sont cependant les seuls masques à améliorer la méthode initiale (DS sans masque) quelle que soit la mesure. De plus, le masque DCT [HPG99] présente les meilleurs résultats pour les mesures habituellement jugées les plus pertinentes (Watson et Komparator). Les masques fondés sur l'interpolation doivent donc être améliorés, ce qui est possible en variant la technique d'interpolation. On remarquera que les mesures "spatiales" (SSIM, CG, C4) avantagent les masques spatiaux par rapport aux masques fréquentiels et inversement. La présence de ces quelques contradictions entre les résultats des différentes mesures encourage la mise en œuvre d'un protocole d'évaluation subjective, ainsi que d'études comme celle de [MACC07].

L'annexe C.1 fournit un exemple d'étude subjective de l'imperceptibilité sur un détail de l'image Lena présentant des zones planes et des transitions abruptes, difficiles à tatouer. Les techniques d'interpolation (en particulier la séquence sous-échantillonnage / suréchantillonnage) peuvent créer des effets de flou au niveau des contours d'une image. Ces artefacts échappent souvent aux mesures de qualité perceptuelle objectives et doivent être soumis à une étude subjective. Cependant, lorsque le masque sert à pondérer une séquence pseudo-aléatoire, les changements du signe de l'erreur d'interpolation éliminent ces artefacts. Il s'agit d'une propriété classique des masques perceptuels en tatouage, par exemple lors de l'utilisation du masque laplacien. Cependant, les techniques d'interpolation basiques telles que l'interpolation bilinéaire peuvent créer un effet de pixellisation sur les contours. Pour résoudre ce problème, des techniques d'interpolation adaptative préservant les contours pourraient être utilisées. Par exemple [GEVS01], qui propose de modifier l'interpolation par B-splines lors de changements de taille d'une image, conviendrait au masque scaling.

Implantations et robustesse

Comme pour les masques classiques présentés dans le paragraphe 1.5.3, plusieurs implantations sont possibles ($\Psi = |h_\Psi * \mathbf{x}|$ et décodage par corrélation par $\mathbf{c}$, $\Psi = h_\Psi * \mathbf{x}$ et décodage par corrélation par $\mathbf{c}/\Psi'$, $\Psi = h_\Psi * \mathbf{x}$ et décodage par corrélation par $\Psi'\mathbf{c}$, où $\Psi' = h_\Psi * \mathbf{z}$). Nous conseillons cette dernière technique, qui offre une amélioration de la robustesse au bruit de l'hôte par rapport à DS sans masque.

	SSIM[1⋆]	$D_W^{2\star}$	$D_{KL}^{2\star}$	CG[2†]	C4[1†]	Komparator[2†]
DS	**0.9870**	**365**	**1370**	**492**	**0.928**	**618**
DS+NVF	0.9944	317	709	445	0.929	585
DS+masque DCT	0.9953	**6**	295	391	**0.944**	**425**
DS+Laplace	0.9978	345	264	371	0.93	733
DS+Alvarez	0.9978	341	331	371	0.934	641
DS+masque bilin	0.9978	347	255	361	0.931	763
DS+masque bspline	0.9975	350	259	347	0.934	788
DS+masque cspline	0.9977	351	246	345	0.936	729
DS+masque scaling	**0.9981**	340	**173**	**259**	0.937	730
SCS	0.9892	497	1693			
ST-SCS	0.987	549	1074			
ST-SCS+NVF	0.9946	454	425			
ST-SCS+Laplace	0.9978	542	189			
ST-SCS+Alvarez	0.9979	506	221			
ST-SCS+masque bilin	0.9978	539	181			
ST-SCS+masque bspline	0.9975	532	173			
ST-SCS+masque cspline	0.9977	530	165			
ST-SCS+masque scaling	0.9981	553	123			
RDM Minkowski, $N_v = 2$	0.987	364	1560			
RDM moy. locale, $N_v = 10$	0.9873	367	1516			
RDM moy. locale, $N_v = 10$	0.9870	359	1553			

[1] valeur optimale : 1, [2] val. opt. : 0

⋆ mesures réalisées pour PSNR=43.5 dB, † pour PSNR=35 dB

3.4 Conclusion

Dans ce chapitre, nous avons mis en évidence les liens entre tatouage et interpolation. A partir du constat des bonnes propriétés perceptuelles de l'interpolation, nous avons proposé des masques perceptuels proportionnels à une erreur d'interpolation. L'erreur d'interpolation est générée heuristiquement, par exemple à l'aide de transformations géométriques. Les masques proposés agissent comme une sorte de détecteur de singularité dans le domaine spatial. Comme la plupart des masques spatiaux existants, ils mettent en valeur les contours et les textures. L'avantage des masques proposés est que leur conception repose sur les techniques d'interpolation déjà existantes.

Les performances perceptuelles doivent être évaluées en pratique. Nous avons constaté que les mesures perceptuelles objectives ne sont pas totalement satisfaisantes. Or est difficile d'effectuer une étude subjective des masques proposés. Cependant, au-delà des implantations proposées, le principe que nous suggérons peut être appliqué à l'ensemble des techniques d'interpolation existantes. Nous espérons donc que des études subjectives pourraient identifier des techniques d'interpolation offrant les mêmes garanties perceptuelles que les masques classiques utilisés en tatouage.

Les masques perceptuels proposés ici seront modulés par un pseudo-bruit. L'image tatouée ne tendra donc pas, même à DWR faible, vers le résultat exact d'une interpolation. Dans le chapitre suivant, nous exploiterons une intuition difficile à valider en pratique autrement que par une étude subjective (cf. annexe C.1) : il est préférable

d'insérer un tatouage directement proportionnel à l'erreur d'interpolation, sans changement de signe. La modulation étant un élément essentiel des techniques de tatouage classiques, nous devons imaginer de nouvelles structures d'insertion et de détection.

Bibliographie

[ARPG02] M. Alvarez-Rodríguez and F. Pérez-González. Analysis of pilot-based synchronization algorithms for watermarking of still images. *Signal Processing : Image Communication*, 17(8) :611–633, 2002.

[BCM00] P. Bas, J-M. Chassery, and B. Macq. Robust watermarking based on the warping of pre-defined triangular patterns. *Proc of SPIE Elec. Imaging, Security and Wat. of Multimedia Content II*, 18 :99–109, 2000.

[BFM05] G. Boato, C. Fontanari, and F. Melgani. Hierarchical deterministic image watermarking via polynomial interpolation. *Proc. of ICIP*, 2005.

[Boo89] F.L. Bookstein. Principal Warps : Thin-Plate Splines and the Decomposition of Deformations. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 11(6) :567–585, 1989.

[Che] CheckMark. http ://watermarking.unige.ch/Checkmark/.

[Cit] City University of Hong Kong Corel Image Database. http ://abacus.ee.cityu.edu.hk/ benjiman/corel_1/.

[CP95] J.-P. Cocquerez and S. Philipp. *Analyse d'images : filtrage et segmentation*. Masson, 1995.

[CR74] E. Catmull and R. Rom. A class of local interpolating splines. *Computer Aided Geometric Design (Proc. of International Conference on Computer Aided Geometric Design '74)*, pages 317–326, 1974.

[DBG+05] P. Dong, J.G. Brankov, N.P. Galatsanos, Y. Yang, and F. Davoine. Digital watermarking robust to geometric distortions. *IEEE Trans. on Image Proc.*, 14(12) :2140–2150, 2005.

[Dod97] N.A. Dodgson. Quadratic interpolation for image resampling. *IEEE. Trans. on Image Processing*, 6 :1322–1326, 1997.

[DP04] F. Davoine and S. Pateux. *Tatouage de documents audiovisuels numériques*. Hermes Science, 2004.

[GEVS01] A. Gotchev, K. Egiazarian, J. Vesma, and T. Saramaki. Edge-preserving image resizing using modified B-splines. *Proc. of ICASSP*, 3 :1865–1868, 2001.

[Gre02] M.L. Green. Statistics of images, the TV algorithm of Rudin-Osher-Fatemi for image denoising and an improved denoising algorithm. *CAM reports, Univ. California, Los Angeles [Online] : http ://www.math.ucla.edu/applied/cam/index.html*, 2002.

[HJMM03] J. Herrera-Joancomarti, J. Minguillon, and D. Megias. A family of image watermarking schemes based on lossy compression. *Proc. of Int. Conf. on Information Technology : Coding and Computing*, pages 559–563, 2003.

[HM99] J. Huang and D. Mumford. Statistics of natural images and models. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'99)*, 1 :541–547, 1999.

[HPG99] J.R. Hernández and F. Pérez-González. Statistical analysis of watermarking schemes for copyright protection of images. *IEEE Proc., Special Issue on Identification and Protection of Multimedia Information*, 87(7) :1142–1166, 1999.

[JP98] J.J.K. Ó Ruanaith and T. Pun. Rotation, scale and translation invariant spread spectrum digital image watermarking. *Signal Proc.*, 66(3) :303–317, 1998.

[Key81] R.G. Keys. Cubic convolution interpolation for digital images. *Transactions on Acoustics, Speech and Signal Processing*, 29 :1153–1160, 1981.

[KJ99] T. Kalker and A. Janssen. Analysis of SPOMF detection. *Proc. of IEEE conference on ICIP*, 1 :316–319, 1999.

[KS95] E. Karabassis and M.E. Spetsakis. An Analysis of Image Interpolation, Differentiation, and Reduction Using Local Polynomial Fits. *CVGIP : Graphical Model and Image Processing*, 57(3) :183–196, 1995.

[LMH01] A.B. Lee, D. Mumford, and J. Huang. Occlusion Models for Natural Images : A Statistical Study of a Scale-Invariant Dead Leaves Model. *Int. Journal of Computer Vision*, 41(1-2) :35–59, 2001.

[LO01] X. Li and M. Orchard. New edge directed interpolation. *IEEE Trans. Image Processing*, 10(10) :1521–1527, 2001.

[LWS97] S. Lee, G. Wolberg, and S. Y. Shin. Scattered data interpolation with multilevel B-splines. *IEEE Transactions on Visualization and Computer Graphics*, 3(3) :228–244, 1997.

[MACC07] E. Marini, F. Autrusseau, P. Le Callet, and P. Campisi. Evaluation of standard watermarking techniques. *SPIE Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents IX*, 2007.

[OMA99] R. Ohbuchi, H. Masuda, and M. Aono. A Shape-Preserving Data Embedding Algorithm for NURBS Curves and Surfaces. *Proc. of the Computer Graphics International (CGI)*, pages 170–177, 1999.

[PF05] A.C. Popescu and H. Farid. Exposing digital forgeries in color filter array interpolated images. *IEEE Trans. on Signal Processing*, 53(10) :3948 – 3959, 2005.

[PLK+06] S.W. Park, L. Linsen, O. Kreylos, J.D. Owens, and B. Hamann. Discrete Sibson Interpolation. *IEEE Transactions on Visualization and Computer Graphics*, 12(2) :243–253, 2006.

[Sib81] R. Sibson. *A Brief Description of Natural Neighbor Interpolation, in Interpreting Multivariate Data*, pages 21–36. V. Barnett, John Wiley and Sons, 1981.

[SKH01] K. Su, D. Kundur, and D. Hatzinakos. A content-dependent spatially localized video watermarked for resistance to collusion and interpolation attacks. *Proc. IEEE Int. Conf. on Image Processing*, 2001.

[SKH02] K. Su, D. Kundur, and D. Hatzinakos. A novel approach to collusion-resistant video watermarking. *Proc. SPIE, Security and Watermarking of Multimedia Contents IV*, 4675, 2002.

[TBU00] P. Thévenaz, T. Blu, and M. Unser. Image interpolation and resampling. In I. Bankman, editor, *Handbook of Medical Imaging, Processing and Analysis*, chapter 25, pages 393–420. Acad. Press, San Diego, USA, 2000.

[Uns99] M. Unser. Splines : A perfect fit for signal and image processing. *IEEE Signal Processing Magazine*, 16(6) :22–38, 1999.

[VHBP99] S. Voloshynovskiy, A. Herrigel, N. Baumgartner, and T. Pun. A stochastic approach to content adaptive digital image watermarking. *International Workshop on Information Hiding*, pages 212–236, 1999.

[Yar97] L.P. Yaroslavsky. Efficient algorithm for discrete sinc interpolation. *Applied Optics*, 36(2) :460–463, 1997.

Chapitre 4

Tatouage substitutif utilisant l'interpolation : W-interp

Sommaire

Notre démarche dans ce chapitre consiste à utiliser l'interpolation comme élément constitutif d'une technique de tatouage. Comme l'interpolation se fait à partir d'échantillons connus du signal, il a été décidé de laisser certains points inchangés (non tatoués). Ils sont utilisés pour calculer des valeurs interpolées, qui constituent le tatouage. L'erreur d'interpolation par rapport aux points originaux est positive ou négative, de moyenne nulle. Cependant, son signe dépend du signal hôte et non d'une clé secrète ou du message. Pour cette raison, nous avons choisi de ne pas nous appuyer sur une technique de tatouage additif pur. De plus, le tatouage ainsi généré est dépendant du document hôte. Il est donc difficile d'envisager une méthode utilisant un dictionnaire de mots de codes prédéfinis. Pour l'ensemble de ces raisons, la technique construite est de type substitutif par imposition de contrainte sur l'hôte (cf. paragraphe 1.2.7). Nous proposons dans la suite une classe d'algorithmes de tatouage, dont nous étudions les propriétés d'imperceptibilité et les performances théoriques. Dans le cas particulier de l'application à l'image, nous étudions la robustesse des algorithmes et construisons un décodeur optimal prenant en compte la distribution de l'erreur d'interpolation. Enfin, le niveau de sécurité de la technique est étudié et des algorithmes pratiques d'attaque sur la sécurité utilisant un algorithme d'Estimation-Maximisation sont proposés. Ce chapitre fera parfois référence à l'annexe C, qui rassemble des développements et des variantes moins génériques de l'algorithme.

4.1 Algorithme W-interp

4.1.1 Principe général

La *fig.*4.1 présente le schéma général de la classe W-interp proposée.
Deux ensembles d'échantillons sont sélectionnés dans le signal $\mathbf{x}$, de coordonnées respectives $\mathcal{G}$ et $\mathcal{S}$. Le tatouage est inséré dans $\mathcal{S} \subset \{1,\dots,N\} \setminus \mathcal{G}$. Soit $N_{\mathcal{S}}$ le cardinal de $\mathcal{S}$ et $P_{\mathcal{S}} = N_{\mathcal{S}}/L$ la redondance. $\mathcal{S}$ est divisé aléatoirement en L sous-ensembles disjoints de taille $P_{\mathcal{S}}$: $\mathcal{S} = \mathcal{S}_1 \cup \cdots \cup \mathcal{S}_L, \forall i \neq j \quad \mathcal{S}_i \cap \mathcal{S}_j = \emptyset$. $\mathcal{S}_l$ est associé au bit m_l du message. Il s'agit donc d'une "mise en forme aléatoire". Cependant, le signal $\mathbf{b}$ mis en forme de taille N n'est pas antipodal : $b_k \in \{0,1\}$. Soit $\mathbf{x}_{|\mathcal{G}}$ la restriction de $\mathbf{x}$ à $\mathcal{G}$. Enfin, soit $\mathbf{g} = \{g^k, k \in \mathcal{S}\}$ un ensemble de fonctions de domaine de définition :

$$g^k : \mathbb{R}^{N_v} \longrightarrow \mathbb{R}$$

où N_v représente la dimension du support de g^k. $\mathbf{g}$ garantit le respect de la contrainte d'imperceptibilité.

Soit $\underline{x}_k$ le vecteur des N_v échantillons de $\mathbf{x}$ situés sur $\mathcal{G}$ dont les coordonnées sont les plus proches de x_k. $\underline{x}_k$ est le vecteur des "voisins" de x_k utilisés comme variable de g^k. A l'insertion, on utilise un signal auxiliaire $g(\underline{\mathbf{x}})$ tel que $g^k(\underline{x}_k) - x_k$ soit faible perceptuellement. Notons que $x_k \notin \underline{x}_k$ n'est pas fourni à g^k. g^k estime des échantillons manquants à partir d'un sous-ensemble de $\mathbf{x}$. Par conséquent, g^k peut être considérée comme une fonction d'interpolation. On se limitera dans la suite à dériver $\mathbf{g}$ des techniques d'interpolation classiques. Cependant, le principe général reste valable pour toute fonction $\mathbf{g}$ respectant les contraintes ci-dessus. $\mathbf{g}$ peut donc être considérée comme une "fonction de similarité". La clé secrète de l'algorithme est $\mathbf{k} = \{\{\mathcal{S}_l\}_{l=1,\dots,L}, \mathbf{g}\}$.

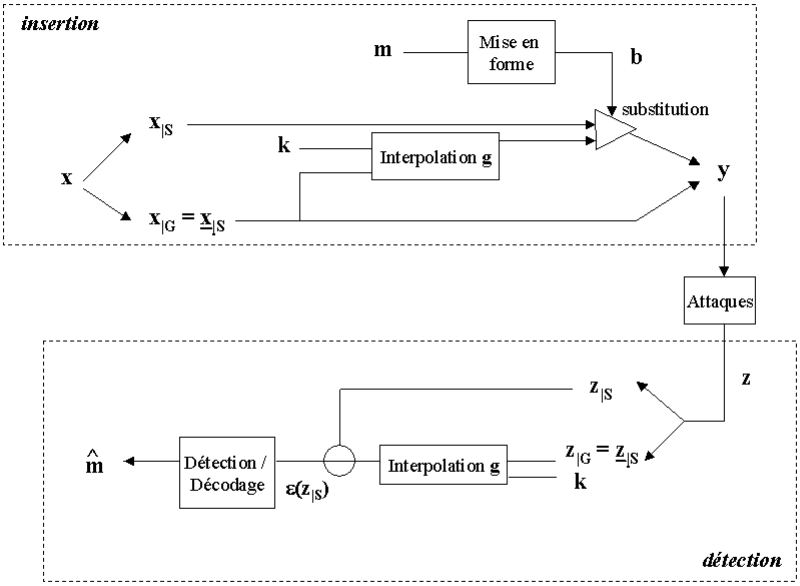

FIG. 4.1 – Classe de méthodes de tatouages W-interp

Insertion : W-interp utilise un **codage par répétition**. On étudie donc l'insertion sur l'élément $k \in \mathcal{S}_l$. Si $m_l = +1$, la valeur x_k est substituée par la valeur correspondante $y_k = g^k(\underline{x}_k)$ fournie par g. Si $m_l = -1$, $y_l = x_k$. L'insertion est donc :

$$y_k = x_k + b_k(g^k(\underline{x}_k) - x_k)$$

Décodage : au décodage, on compare $\mathbf{z}_{|\mathcal{S}_l}$ et $g(\underline{\mathbf{z}}_{|\mathcal{S}_l})$. Soit $\mathbf{r} = g(\underline{\mathbf{z}}_{|\mathcal{S}_l}) - \mathbf{z}_{|\mathcal{S}_l}$ le résultat de cette comparaison. Pour un bit m_l donné, on compare l'erreur au sens des moindres carrés

$$\rho_l^2 \triangleq \frac{1}{|\mathcal{S}_l|} \sum_{k \in \mathcal{S}_l} r_k^2$$

à un seuil η dépendant du signal. Si $\rho_l^2 < \eta$, la décision est $d_l = +1$ (l'image reçue est proche de l'image interpolée), sinon $d_l = -1$. η peut être choisi empiriquement comme la moyenne des résultats des décodages :

$$\eta = \frac{1}{L} \sum_{l=1}^{L} \rho_l^2$$

On peut calculer un seuil théorique plus performant sous des hypothèses adéquates sur la distribution de $g(\underline{x}_k) - x_k$ (cf. paragraphe 4.2). Une amélioration du seuil empirique utilisant un décodage itératif est également proposée dans l'annexe C.2.

Ce cadre général fournit des algorithmes de tatouage **aveugles**, car $\mathbf{x}$ n'est pas utilisée au décodage. W-interp est un algorithme de tatouage substitutif et à état de l'hôte

connu, car l'insertion prend en compte la valeur de $\mathbf{x}$. W-interp est une méthode à **rejet des interférences de l'hôte** car en l'absence d'attaque, on obtient un décodage parfait. Cette propriété n'est cependant vérifiée que dans le cas d'un tatouage bas débit (cf. paragraphe 4.6.1). On peut alors insérer jusqu'à $N_{\mathcal{S}}$ bits. La redondance d'insertion est $P_{\mathcal{S}}$, mais le débit reste défini par $R = 1/P$: on prend en compte la grille d'interpolation. Le débit accessible est $N_{\mathcal{S}}/N$. De plus, W-interp est un algorithme de tatouage informé. En effet, il utilise $\mathbf{x}$ pour générer $\mathbf{w}$, dans le souci de respecter un modèle perceptuel, donc il s'agit de **codage informé**. Par contre, la seule stratégie d'insertion informée est ici le rejet des interférences de l'hôte, *i.e.* maximiser la détection à distorsion constante et en l'absence d'attaque. Une extension de W-interp à l'insertion informée devrait utiliser la connaissance du décodeur à l'insertion pour mettre en place une stratégie donnée (cf. paragraphe 4.3.3). La méthode proposée a l'intérêt de pouvoir être étudiée de manière théorique, ce qui sera le cas dans le paragraphe 4.2. Du fait de la substitution, le tatouage multiple n'est possible que si $N_{\mathcal{S}}$ est faible et si chaque utilisateur tatoue un partie distincte de $\{1, \ldots, N\} \setminus \mathcal{G}$ (division spatiale, la division fréquentielle ou par code étant impossible). Un algorithme donné est caractérisé par le choix d'un ensemble de fonctions de similarité $\mathbf{g}$, d'une grille $\mathcal{G}$ et des positions $\mathcal{S}$ des points à tatouer. Des exemples d'application de W-interp au tatouage d'images seront donnés dans le paragraphe 4.4.

4.1.2 Analyse de l'algorithme

Scenario d'application : l'utilisation du codage informé impose un scénario d'application où $\mathbf{x}$ est connu à l'insertion. Ceci inclut par exemple la protection des droits d'auteurs. De plus, l'insertion est pour l'instant effectuée dans le domaine spatial, pour des raisons perceptuelles. Pour les algorithmes classiques, un domaine transformé est pourtant souvent préféré dans la plupart des scénarios d'applications, afin d'améliorer la robustesse. Il serait donc intéressant d'identifier des applications où le calcul d'un domaine transformé est impossible (pour des raisons de complexité par exemple).

Sécurité : dans QIM et DS, la sécurité repose sur la modulation de $\mathbf{b}$ ou des quantificateurs par un code. S'il ignore le code, l'attaquant ne peut pas différencier les deux centroïdes. Pour W-interp, si l'attaquant connaît $\mathbf{g}$, il connaît également $\mathbf{b}$. De plus, une modulation de $\mathbf{b}$ n'est pas possible car on ne peut pas combiner les résultat de décodages indépendants comme pour QIM ou DS : la redondance fait partie intégrante du décodage. On fait donc varier ici les coefficients du filtre d'interpolation $\mathbf{g}$. De plus, $\mathcal{S}$ est généré aléatoirement. L'influence et le choix des paramètres de sécurité dans $\mathbf{g}$ seront étudiés au paragraphe 4.7.

Puissance d'insertion : considérons que les bits $\{-1, +1\}$ sont équiprobables. Alors l'expression théorique du rapport document à tatouage DWR est

$$\text{DWR} = \frac{2\sigma_{\mathbf{x}}^2 N}{\sigma_{\epsilon(\mathbf{x})}^2 N_{\mathcal{S}}} \tag{4.1}$$

où $N_{\mathcal{S}}/2$ est le nombre de points modifiés. En pratique, on fixe $N_{\mathcal{S}}$ en connaissant $\sigma_{\epsilon(\mathbf{x})}^2$ et le DWR désiré. Enfin, il est important de remarquer que la puissance d'insertion maximale possible est limitée. En effet, $N_{\mathcal{S}} < N/2$. Donc DWR$> \frac{4\sigma_{\mathbf{x}}^2}{\sigma_{\epsilon(\mathbf{x})}^2}$.

Fonction d'interpolation : l'imperceptibilité sera étudiée au paragraphe 4.5. Notons que g^k sera souvent linéaire. g^k agira donc comme un filtre local. La condition d'imperceptibilité impose que $\mathbf{w}$ modifie les hautes et moyennes fréquences de $\mathbf{x}$. g^k agit donc comme un filtre passe-bas, et la méthode de tatouage consiste en une modification des coefficients passe-haut de $\mathbf{x}$. De plus, lorsque g^k est linéaire, une propriété importante est que $\forall \lambda \in \mathbb{R}, g(\lambda \underline{x}_k) = \lambda g(\underline{x}_k)$.

Quantification du document : en pratique, $\mathbf{y}$ est quantifié en N_L niveaux de luminance à l'insertion ($N_L = 256$ par exemple pour une image de 8 bits). Le tatouage inséré est donc issu d'une version quantifiée de l'erreur d'interpolation. On peut inclure cette quantification dans g_k, qui devient alors non linéaire. La technique est toujours à rejet des interférences de l'hôte : on compare à la réception avec la version quantifiée. Son impact sur le décodage est étudié dans le paragraphe 4.2, et son rôle dans la sécurité est présenté dans le paragraphe 4.7.

4.2 Performances théoriques face au bruit additif gaussien

Dans ce paragraphe, on calcule les performances théoriques de détection et de décodage de W-interp face à l'attaque d'ajout de bruit blanc gaussien $\mathbf{n}$ de variance $\sigma^2_{\mathbf{n}}$. Cette étude correspond à deux scénarios. Dans le premier, $\sigma^2_{\mathbf{n}}$ est connue au décodage. L'étude théorique permet alors de calculer un seuil de détection optimal, qui améliore les performances par rapport au seuil empirique. Si $\sigma^2_{\mathbf{n}}$ n'est pas connu, le seuil empirique est utilisé en pratique mais l'étude théorique permet d'obtenir une borne supérieure des performances de W-interp. La validité des calculs est soumise à l'hypothèse que la distribution de $g(\underline{\mathbf{x}}) - \mathbf{x}$ est gaussienne, et au fait que les g^k soient linéaires.

4.2.1 Influence du bruit gaussien

Soit $\underline{x}_{k,j}$ l'élément j du vecteur $\underline{x}_k$ et g^k_j le poids de $\underline{x}_{k,j}$ dans g^k. Rappelons que $\sum_{j=1}^{N_v} g^k_j = 1$. Le résultat de la comparaison au point k est :

$$r_k = g^k(\underline{z}_k) - z_k = \sum_{j=1}^{N_v} g^k_j(\underline{y}_{k,j} - y_k) \sum_{j=1}^{N_v} g^k_j(\underline{n}_{k,j} - n_k)$$

Soit $\epsilon(x) = g(\underline{x}) - x$. Alors

$$r_k = \epsilon(x_k) + \epsilon(n_k),$$

$\epsilon(\mathbf{x})$ et $\epsilon(\mathbf{n})$ étant les contributions respectives à $\mathbf{r}$ de l'image et du bruit. Comme $\mathrm{E}[\mathbf{n}] = 0$, $\mathrm{E}[\epsilon(\mathbf{n})] = 0$ et comme les échantillons n_k sont indépendants, de moyenne nulle et de même variance,

$$\begin{aligned}\mathrm{Var}[\epsilon(\mathbf{n})] &= \mathrm{E}[n_k^2 + \sum_{j=1}^{N_v} {g^k_j}^2 \underline{n}^2_{k,j}] \\ &= \mathrm{E}[\mathbf{n}^2]\mathrm{E}[1 + \sum_{j=1}^{N_v} {g^k_j}^2]\end{aligned}$$

Si $\Delta \triangleq \mathrm{E}[\sum_{j=1}^{N_v} {g_j^k}^2]$, la variance de $\epsilon(\mathbf{n})$ peut donc s'exprimer sous la forme

$$\sigma^2_{\epsilon(\mathbf{n})} = (1+\Delta)\sigma^2_{\mathbf{n}}$$

En particulier, pour tout g_j^k de moyenne $\frac{1}{N_v}$ et de variance $\sigma^2_{\mathbf{g}}$,

$$\Delta = N_v(\sigma^2_{\mathbf{g}} + \frac{1}{N_v^2}) \tag{4.2}$$

Le choix de la distribution de g_j^k résulte d'un compromis entre performance et imperceptibilité d'une part, et sécurité de l'autre. En effet, plus la variance de g_j^k est faible, plus on est proche d'une interpolation idéale, mais il y a peu de différence entre deux tatouages avec des paramètres g_j^k différents.

4.2.2 Détecteur sous-optimal : hypothèse gaussienne

Dans cette section, l'erreur d'interpolation est modélisée comme une variable gaussienne de moyenne nulle et de variance $\sigma^2_{\epsilon(\mathbf{x})}$.

Problème de la détection

Pour le problème de la détection, on se placera pour simplifier et sans perte de généralité dans le cas d'un tatouage d'un seul bit ($L = 1$) avec $m_l = 1$. Dans le cas de plusieurs bits équiprobables, il faut cependant remplacer $\sigma^2_{\epsilon(\mathbf{x})})$ par $\sigma^2_{\epsilon(\mathbf{x})})/2$ dans la suite. Le test s'effectue entre deux hypothèses :
Hypothèse H_1 : présence d'un tatouage

$$r_k = g(\underline{z}_k) - z_k = g(\underline{y}_k) - g(\underline{x}_k) + \sum_{j=1}^{N_v} g_j^k(\underline{n}_{k,j} - n_k) = \epsilon(n_k)$$

Hypothèse H_0 : absence de tatouage

$$r_k = g(\underline{z}_k) - z_k = \epsilon(x_k) + \epsilon(n_k)$$

$\epsilon(x_k)$ gaussien, on a : $R \sim \mathcal{N}(0, (1+\Delta)\sigma^2_{\mathbf{n}} + \sigma^2_{\epsilon(\mathbf{x})})$. La statistique de test de Neyman-Pearson est donc :

$$\ln\left(\frac{\prod_k \frac{1}{\sqrt{2\pi(1+\Delta)\sigma^2_{\mathbf{n}}}} e^{-\frac{r_k^2}{2(1+\Delta)\sigma^2_{\mathbf{n}}}}}{\prod_k \frac{1}{\sqrt{2\pi(1+\Delta)\sigma^2_{\mathbf{n}}+\sigma^2_{\epsilon(\mathbf{x})}}} e^{-\frac{r_k^2}{2(1+\Delta)\sigma^2_{\mathbf{n}}+\sigma^2_{\epsilon(\mathbf{x})}}}}\right)$$
$$= P_{\mathcal{S}} \ln\left(\frac{(1+\Delta)\sigma^2_{\mathbf{n}} + \sigma^2_{\epsilon(\mathbf{x})}}{(1+\Delta)\sigma^2_{\mathbf{n}}}\right) - \frac{1}{2}\left(\frac{1}{(1+\Delta)\sigma^2_{\mathbf{n}}} - \frac{1}{(1+\Delta)\sigma^2_{\mathbf{n}} + \sigma^2_{\epsilon(\mathbf{x})}}\right)\sum_k r_k^2$$

Ce qui revient à une statistique de test

$$\mathbf{T} = \sum_{\mathcal{S}} r_k^2 \lessgtr \eta$$

Comme r_k est gaussien, $\mathbf{T}$ suit une loi du Chi2 à P degrés de liberté χ^2_P. Pour P grand, le théorème Central-Limite permettrait une approximation gaussienne de la distribution de la statistique de test. L'un des intérêts serait de pouvoir résoudre explicitement $f_{T|H_1}(\eta) - f_{T|H_{-1}}(\eta) = 0$, qui se ramènerait facilement à la racine d'un polynôme du second ordre (cf. paragraphe 4.4.4). Toutefois, le compromis imperceptibilité/capacité conduit à de petites valeurs de P. Les résultats en pratique montrent que le seuil est évalué de manière plus précise avec une distribution χ^2_P.

Alors à P_{fa} donnée,

$$\eta = (1+\Delta)\sigma^2_{\mathbf{n}} \mathrm{F}^{-1}_{\chi^2_P}(1 - \mathrm{P}_{\text{fa}})$$

et

$$P_{\text{nd}} = \mathrm{F}_{\chi^2_P}\left(\frac{\eta}{(1+\Delta)\sigma^2_{\mathbf{n}} + \sigma^2_{\epsilon(\mathbf{x})}}\right)$$

De plus, cette valeur du seuil correspond à $P_{\text{fa}} = 1 - \mathrm{F}_{\chi^2_P}\left(\frac{\eta}{(1+\Delta)\sigma^2_{\mathbf{n}}}\right)$.

Les résultats expérimentaux dans l'application W-bilin (cf. paragraphe 4.4) de la courbe COR de la *fig.* 4.2 sont en bonne adéquation avec la théorie.

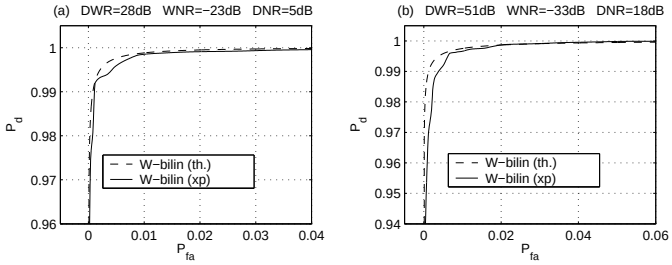

FIG. 4.2 – COR théorique et pratique, W-bilin

Problème du décodage

Dans le problème du décodage, $L > 1$ et $m_l \in \{-1, 1\}$. Pour un bit l et un ensemble de points $\mathcal{S}_l$ donnés, les calculs effectués pour la détection sont toujours valables, mais désormais on travaille sur les hypothèses
H_1 : le bit inséré est $m_l = 1$
H_{-1} : le bit inséré est $m_l = -1$.
On a vu dans le cas du tatouage par étalement de spectre que les problèmes de la détection et du décodage nécessitent un traitement distinct. Notamment, pour DS et PCC, on a dû recourir à un détecteur multi-bit sous-optimal (cf. paragraphe 2.1.3). Cependant, dans le cas particulier de W-interp, on passe aisément de la détection au décodage. Sa particularité est en effet que l'hypothèse (H_0 : tatouage absent) et l'hypothèse (H_{-1} : le bit inséré est $m_l = -1$) se confondent sur un support $\mathcal{S}_l$ donné.

Alors la performance est évaluée par la probabilité d'erreur :

$$\text{TEB} = \mathrm{p}[m_l = 1]\mathrm{p}[T > \eta_{\text{th}}|H_1] + \mathrm{p}[m_l = -1]\mathrm{p}[T < \eta_{\text{th}}|H_0]$$

Le seuil η_{th} est choisi pour minimiser le TEB dans le cas de bits -1 et +1 équiprobables, et doit donc vérifier $\frac{\partial \text{TEB}}{\partial \eta}_{|\eta=\eta_{\text{th}}} = 0$, soit $f_{T|H_1}(\eta_{\text{th}}) - f_{T|H_0}(\eta_{\text{th}}) = 0$. Cela correspond à

$$\frac{1}{(1+\Delta)\sigma^2_{\mathbf{n}} + \sigma^2_{\epsilon(\mathbf{x})}} \mathrm{f}_{\chi^2_P}\left(\frac{\eta_{\text{th}}}{((1+\Delta)\sigma^2_{\mathbf{n}} + \sigma^2_{\epsilon(\mathbf{x})}}\right) = \frac{1}{(1+\Delta)\sigma^2_{\mathbf{n}}} \mathrm{f}_{\chi^2_P}\left(\frac{\eta_{\text{th}}}{(1+\Delta)\sigma^2_{\mathbf{n}}}\right)$$

Donc

$$\text{TEB} = \frac{1}{2}(\text{F}_{\chi^2_P}\left(\frac{\eta_{\text{th}}}{(1+\Delta)\sigma^2_{\mathbf{n}} + \sigma^2_{\epsilon(\mathbf{x})}}\right) + 1 - \text{F}_{\chi^2_P}\left(\frac{\eta_{\text{th}}}{(1+\Delta)\sigma^2_{\mathbf{n}}}\right)$$

avec η_{th} calculé numériquement. Sur l'exemple de la *fig.* 4.3, on choisirait $\eta_{\text{th}} = 11750$. La *fig.* 4.4 montre que ce TEB théorique en fonction de η est proche du TEB empirique dans l'application W-bilin décrite dans la suite, et que le seuil optimal permet d'améliorer le TEB par rapport au seuil empirique.

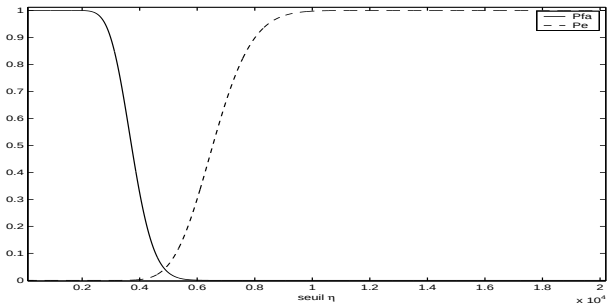

FIG. 4.3 – Probabilités d'erreur en décidant H_1 ou H_0 (P=150, $\sigma^2_{\mathbf{n}} = 40$, W-interp, Lena)

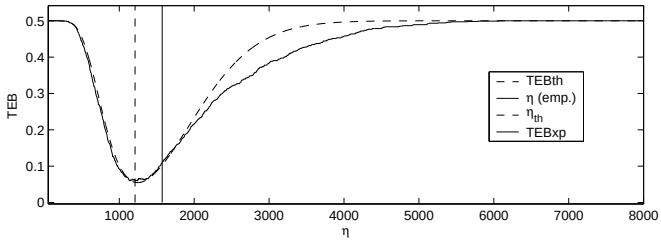

FIG. 4.4 – Choix du seuil : théorique, empirique, L=1024, DWR=28 dB, WNR=-10dB, W-interp, Bateaux

Performances théoriques face à l'attaque de gain

Dans l'attaque de gain, $\mathbf{z} = \rho(\mathbf{y} + \mathbf{n})$. W-interp est invariant au facteur d'échelle grâce à la linéarité de $\mathbf{g}$:

$$\hat{m}_l = \rho^2 ||\mathbf{z} - Q_{\mathbf{b}}(\mathbf{z})||^2_l \lesseqgtr \eta$$

et η est fixé de façon adaptative. Si $\sigma^2_{\mathbf{n}} \neq 0$, on se ramène à des bruits de variance $\rho\sigma^2_{\mathbf{n}}$ contre $\rho\sigma^2_{\epsilon(\mathbf{x})}$ et les performances sont identiques.

4.3 Extension à l'insertion informée

Nous montrons dans cette partie comment, en plus d'être une technique de codage informé, W-interp peut être considérée comme une technique d'insertion informée. Tout d'abord, nous étudions les liens entre W-interp et QIM. Cette analogie nous permet de proposer une adaptation de W-interp à la compensation des distorsions. Enfin, des techniques d'insertion informée utilisant la terminologie de Cox et Miller sont proposées. Dans l'annexe C.3, nous proposons une autre extension de W-interp à l'insertion informée, fondée cette fois-ci sur l'optimisation.

4.3.1 Liens entre W-interp et la catégorisation aléatoire

Lien entre W-interp et RDM

Bien que W-interp ait initialement été construite comme une technique substitutive originale, ses propriétés (tatouage informé, rejet des interférences de l'hôte) rapprochent la méthode des techniques quantificatives. Il est intéressant de reformuler W-interp dans le cadre de QIM et RDM (cf. paragraphe 1.3.2).

Le décodage dans W-interp est fondé sur la relation $g(\underline{y}_k) = g(\underline{x}_k)$, sorte d'idempotence qui est à rapprocher de la relation $Q_{\Delta,b_k\tau_k}(y_k) = Q_{\Delta,b_k\tau_k}(Q_{\Delta,b_k\tau_k}(x_k)) = Q_{\Delta,b_k\tau_k}(x_k)$ qui est la base du décodage des techniques quantificatives. Cependant, le décodage de W-interp n'est sans erreur que pour $b_k = 0$: $y_k - g(\underline{y}_k) = g(\underline{x}_k) - g(\underline{y}_k) = 0$. Pour $b_k = -1$, $y_k - g(\underline{y}_k) = x_k - g(\underline{x}_k) \neq 0$ en moyenne uniquement, car $g(\underline{X}) - X$ suit une GGD. Le rejet des interférences se fait donc en moyenne : même si $P_{\mathcal{S}} > 1$ la probabilité pour que $\sum_{\mathcal{S}_l} Q(\epsilon(x_k))^2 = 0$, où Q est un quantificateur de pas 1, est très faible. Le décodage n'est cependant que sous-optimal dans W-interp, et doit s'appuyer sur une redondance par répétition (cf. paragraphe 4.6.1).

Pour rapprocher W-interp de la formulation de QIM, nous définissons les "quantificateurs linéaires" Q_0 et Q_1 suivants :

$$Q_1(x_k) \triangleq g(\underline{x}_k) \quad \text{et} \quad Q_0(x_k) \triangleq x_k$$

On peut exprimer l'insertion de W-interp comme :

$$y_k = Q_{b_k}(x_k)$$

W-interp utilise donc 2 centroïdes : x_k et $\underline{x}_k$. Le pas de quantification est "uniforme" dans le sens où g est indépendant de l'échantillon k. Ce n'est plus le cas lorsqu'on introduit le secret g^k. Les "quantificateurs" de W-interp ont pour principales propriétés :

- les espaces image de Q_0 et Q_1 ne sont pas disjoints, ni construits en optimisant leur éloignement
- il n'y a pas de symétrie entre les espaces images de Q_0 et Q_1
- les espaces image de Q_0 et Q_1 sont très proches perceptuellement de la variable (c'est l'idée de départ de la méthode)
- Q_0 et Q_1 sont linéaires si g l'est, ce qui est très différent des quantificateurs classiques
- le pas de "quantification" est nul pour Q_0, et **adaptatif** pour Q_1.
- la pas de quantification est inconnu au décodage. Il est recalculé grâce à un signal adjacent (les échantillons de la grille d'interpolation). On doit donc transmettre une information sur le "dictionnaire" en même temps que le tatouage, ce qui diminue le débit accessible.

L'utilisation de fonctions d'encodage Q_0 et Q_1 classe désormais W-interp parmi les techniques QIM. Dans l'article de Chen et Wornell [CW01], les fonctions d'encodage sont des approximations du signal dont l'espace image couvre l'espace des possibles, définition qui convient à Q_0 et Q_1. Cependant, dans [CW01], les auteurs précisent qu'à leur sens les espaces images des fonctions d'encodage doivent "à tout le moins être disjoints", ce qui les conduit à n'utiliser dans la pratique que des quantificateurs classiques, discontinus. Nous montrons avec W-interp qu'une méthode de tatouage peut néanmoins s'appuyer sur des fonctions d'encodage non disjointes et continues.

Le décodage effectue :

$$\hat{m}_l = \|\mathbf{z} - Q_1(\mathbf{z})\|_l^2 \lessgtr \eta$$

où $\|.\|_l$ désigne la norme sur l'ensemble $\mathcal{S}_l$. On ne peut pas s'appuyer sur la structure $\arg\min$ car $Q_0(Q_1(x_k)) = Q_1(x_k)$, ce qui est fondamentalement différent du QIM. Le seuil η doit donc être fixé par un test de Neyman-Pearson, et non par un décodeur de distance minimum. De plus, le "pas de quantification" de Q_1 est inconnu (il n'est connu qu'en moyenne). Il s'agit d'un décodage sous-optimal. L'avantage de cette technique est cependant d'inclure le bruit dans le calcul du seuil de décodage.

La robustesse aux attaques valumétriques de W-interp est proche de celle de RDM : elle vient de la relation $\forall \rho \in \mathbb{R}, g(\rho \underline{x}_k) = \rho g(\underline{x}_k)$. Les deux techniques sont invariantes à un facteur d'échelle. Comme RDM, W-interp n'est pas invariante à une attaque valumétrique non affine comme la correction gamma. D'autre part, si ρ est variable, aucune des deux méthodes n'est totalement invariante à l'attaque. Dans RDM, on a donc un compromis entre une bonne robustesse à une attaque de gain variable (N_v faible) et la robustesse aux autres attaques (N_v grand, grâce à l'approche causale). Dans W-interp, on a le choix de N_v mais il est souvent faible dans les implantations proposées ($N_v = 4$ ou 9) et ne résulte pas d'un compromis, ce qui pourrait être un avantage par rapport à RDM. Une étude expérimentale doit donc être faite. W-interp et RDM partagent le même défaut signalé par [Bas05] : certaines zones de l'image sont plus sensibles que d'autres à un AWGN. Une comparaison entre RDM et W-interp face à diverses attaques valumétriques sur des images naturelles est effectuée dans le paragraphe 4.6.3.

Dans RDM, $\underline{\mathbf{x}}$ est constitué d'éléments de $\mathbf{y}$. En effet, dans RDM chaque échantillon est tatoué séquentiellement en anticipant sur le décodage. Dans W-interp, il n'y a pas de causalité du support d'interpolation. Anticiper le décodage contraindrait donc à résoudre un système linéaire très complexe. C'est pourquoi nous avons fait le choix de ne pas modifier certaines valeurs de $\mathbf{x}$, et $\underline{\mathbf{x}}$ est disjoint de $\mathbf{x}_{|\mathcal{S}}$: on n'a pas à anticiper.

Enfin, il est intéressant de constater que malgré la possibilité d'utiliser un signal d'agitation secret, une piste pour améliorer la sécurité des algorithmes quantificatifs consiste à rendre secrète la fonction g de RDM (cf. paragraphe 1.4.2), choix que nous avons également effectué pour W-interp. Contrairement à W-interp, ce choix rend cependant le contrôle de la distorsion d'insertion difficile pour RDM [PFCTPPG06].

Les principales qualités de W-interp par rapport à QIM ou SCS sont sa robustesse intrinsèque aux attaques valumétriques et un masquage perceptuel intrinsèque. Ces propriétés auraient sans doute pu être obtenues, au prix de complications de la technique, en combinant RDM avec une analyse perceptuelle. Inversement, on peut combiner RDM et un masque perceptuel fondé sur l'erreur d'interpolation (cf. paragraphe 3.3.2). Malgré cette analogie, des choix heuristiques font la singularité de la méthode W-interp.

La démarche heuristique de W-interp, reformulée par rapport à RDM, est la suivante : le point de départ de l'algorithme est de substituer à x_k le résultat d'une interpolation $g^k(\underline{x}_k)$. Ce choix de g est original, car les fonctions utilisées dans RDM sont habituellement des normes ou des moyennes locales et causales (cf. paragraphe 1.3.2). C'est le premier choix heuristique délibéré, qui est justifié par des considérations perceptuelles. C'est ce choix qui impose l'emploi de "quantificateurs linéaires", qui est un élément très original de la méthode. On pourrait alors utiliser 2 grilles d'interpolation de fonctions g distinctes selon la valeur de b_k. La technique se serait alors rapprochée de RDM avec un quantificateur linéaire. Cependant, on a fait un deuxième choix pour éloigner les deux classes et donc renforcer la robustesse : celui de laisser $y_k = x_k$ si $b_k = 0$, troisième élément original. Le fait de faire porter la sécurité par g et le dé-

tecteur utilisé sont ensuite imposés par les choix précédents. Les choix heuristiques se font à l'avantage de l'imperceptibilité, au détriment de la robustesse (due au détecteur imposé). On s'attend donc à de moins bonnes performances au décodage qu'une combinaison RDM+ST+masque perceptuel face aux attaques de gain et AWGN.

Combinaison de W-interp avec une transformée d'étalement

L'utilisation d'une transformée d'étalement permet souvent d'améliorer la robustesse à un très fort bruit additif (exemple : ST-SCS). Grâce à l'étalement, seule la composante du bruit parallèle au code d'étalement nuit au décodage. Cependant, il est impossible de combiner W-interp à une transformée d'étalement à cause de la linéarité de Q_0 et Q_1 lorsque les g^k sont linéaires. En effet,

$$Q_1(\frac{1}{P}\sum c_k x_k) = \frac{1}{P}\sum c_k g(\underline{x}_k) \quad \text{et} \quad Q_0(\frac{1}{P}\sum c_k x_k) = \frac{1}{P}\sum c_k x_k$$

On perd l'effet de masque perceptuel de $\mathbf{g}$, qu'on retrouve habituellement dans la séquence $\mathbf{c}$ dans ST-SCS. La diminution de la puissance du bruit est compensée par la disparition de la redondance au décodage, donc le détecteur sous-optimal de W-interp n'est pas amélioré.

4.3.2 W-interp et compensation des distorsions

Cas général

Étudions désormais le "pas de quantification" de Q_1. On a vu que Δ est relié à g (cf. paragraphe 4.2.1). On dira désormais que $y_k = Q_{\Delta,b_k}(x_k)$. La compensation des distorsions consiste donc, comme pour QIM, à utiliser la grille $Q_{\Delta/\alpha,b_k}$ au lieu de Q_{Δ,b_k}.

Lorsque Δ augmente à N_S constant, la distance entre les distribution de $\mathbf{T}$ sous H_0 et H_1 augmente, mais DWR diminue. On propose donc la stratégie d'insertion suivante sous H_1, avec $\Delta^\alpha > \Delta$:

$$\begin{aligned} y_k &= g(\underline{x}_k) + (1-\alpha)(x_k - g(\underline{x}_k)) \\ &= x_k + \alpha(g(\underline{x}_k) - x_k) \end{aligned}$$

Sous H_1, $R \sim \mathcal{N}(0, (1+\Delta^\alpha)\sigma^2_{\mathbf{n}} + (1-\alpha)^2\sigma^2_{\epsilon(\mathbf{x})})$: la compensation des distorsions ajoute des interférences au décodage. Si $\alpha = 1$, on retrouve le schéma classique. Soient $\mathbf{g}^\alpha$ les fonctions d'interpolation correspondant à Δ^α. Supposons l'influence de $\mathbf{g}^\alpha$ sur $\epsilon(\mathbf{x})$ connue, de variance $\sigma^2_{\epsilon(\mathbf{x})(\mathbf{g}^\alpha)}$. D'après l'équation (4.1), à distorsion constante

$$\alpha = \sqrt{\frac{\sigma^2_{\epsilon(\mathbf{x})}}{\sigma^2_{\epsilon(\mathbf{x})(\mathbf{g}^\alpha)}}}$$

Alors η_{th} et le TEB dépendent des variances $(1+\Delta^\alpha)\sigma^2_{\mathbf{n}}$ et $(1+\Delta^\alpha)\sigma^2_{\mathbf{n}} + (1-\alpha)^2\sigma^2_{\epsilon(\mathbf{x})(\mathbf{g}^\alpha)}$. On peut donc calculer numériquement Δ^* qui minimise ce TEB. En pratique, on peut fournir au décodeur une clé $\{\mathcal{S}, g, \mathcal{T}\}$ avec g une fonction d'interpolation de base de moyenne $\frac{1}{N_v}$ et $\mathcal{T}$ ayant une distribution de moyenne nulle et de variance 1. A l'insertion comme au décodage, d'après l'équation (4.2), il suffira ensuite d'utiliser $\mathbf{g} = g + a\mathcal{T}$ avec $a \triangleq \frac{\Delta^*}{N_v} - \frac{1}{N_v^2}$.

Application à l'image

En l'absence de modèle simple de $\mathbf{x}$, les paramètres $\sigma^2_{\epsilon(\mathbf{x})(\mathbf{g}^\alpha)}$ peuvent être calculés numériquement pour chaque document. Dans ce paragraphe, nous proposons une alternative dans le cas d'une application à l'image. Pour modéliser l'influence de $\mathbf{g}$ sur $\epsilon(\mathbf{x})$, on propose cependant d'utiliser le modèle Markov-Gauss suivant [SMCM05] : la différence $\mathbf{u}$ entre deux pixels voisins de $\mathbf{x}$ est supposée gaussienne centrée : $\mathbf{u} \sim \mathcal{N}(0, \sigma^2_{\mathbf{u}})$. Sous l'hypothèse (abusive) d'indépendance des éléments de $\mathbf{u}$, on peut montrer comme dans l'annexe C.4 pour $\epsilon(\mathbf{n})$ que

$$\sigma^2_{\epsilon(\mathbf{x})(\mathbf{g})} = \Delta\sigma^2_{\mathbf{u}}$$

La validité de ce modèle dépend de $\mathbf{x}$. Notamment, il est bien vérifié par l'image Lena lorsque Δ est faible (cf. *fig.* 4.5).

W-bilin est une implantation particulière de W-interp, présentée dans le paragraphe 4.4.2. Les courbes théoriques des *fig.* 4.6 et 4.7 montrent que selon ce modèle, DC-W-bilin apporte une nette amélioration des performances. Les résultats expérimentaux utilisant Δ^* calculé théoriquement grâce au modèle précédent confirment l'intérêt de DC-W-interp (cf. *fig.* 4.8), même si l'amélioration est moindre.

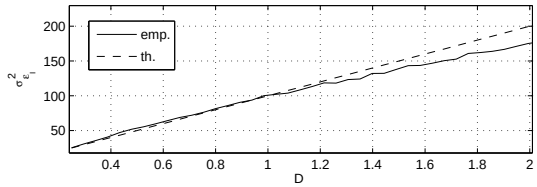

FIG. 4.5 – $\sigma^2_{\epsilon(\mathbf{x})}$ en fonction de Δ, Lena

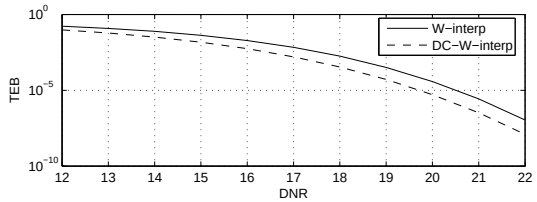

FIG. 4.6 – Amélioration des performances théoriques par DC-W-bilin, Lena, DWR=28 dB, $L = 256$, $P_{\mathcal{S}} = 178$

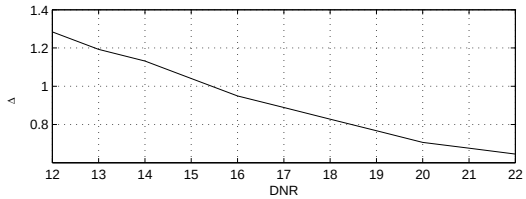

FIG. 4.7 – Choix optimal de Δ pour DC-W-bilin, Lena, DWR=28 dB, $L = 256$, $P_{\mathcal{S}} = 178$

4.3.3 Stratégies d'insertion informée

La stratégie d'insertion classique, utilisée également dans la version de base de W-interp, consiste à rendre maximale la détection à distorsion fixée. Cependant, la

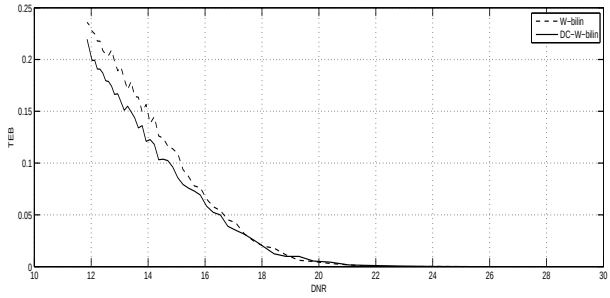

FIG. 4.8 – Amélioration des performances pratiques par DC-W-bilin, Lena, DWR=28 dB, $L = 256$, $P_{\mathcal{S}} = 178$

connaissance du détecteur (et de ses performances théoriques) lors de l'insertion permet d'appliquer d'autres stratégies d'insertion, autour des critères de détection, distorsion et robustesse [MCB00][DFHS03]. La distorsion sera mesurée ici par DWR. On choisit de mesurer la détection par la distance de Kullback-Leibler D_{KL} (cf. paragraphe 4.4.4), dans le cas où $\sigma^2_{\mathbf{n}}$ est nul. La robustesse est un critère à définir pour chaque technique. Pour DS, il s'agit de la puissance $\sigma^2_{\mathbf{n}}$ de bruit qu'un pirate doit ajouter pour fausser le détecteur [MCB00]. Pour W-interp, le seuil η_{th} dépend déjà de $\sigma^2_{\mathbf{n}}$. On préfère donc choisir comme critère de robustesse le TEB, calculé numériquement en fonction des distributions de $\mathbf{T}$ sous H_0 et H_1, à $\sigma^2_{\mathbf{n}}$ connu.

Maximiser la robustesse à distorsion constante

On veut minimiser le TEB à $\sigma^2_{\mathbf{n}}$ donné et à distorsion fixe. Alors DC-W-interp constitue déjà une stratégie d'insertion pratique pour ce problème. Notons que DC-W-interp nuit en revanche à la détection (cf. Fig 4.9 : sans DC, D_{KL} serait infini).

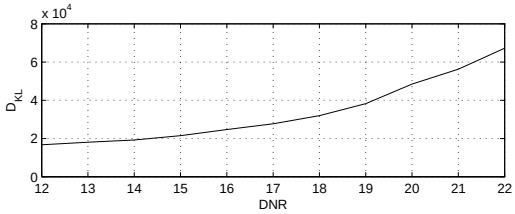

FIG. 4.9 – D_{KL} pour DC-W-bilin si l'attaque n'a pas lieu

Minimiser la distorsion à détection constante

Cette stratégie est inutile pour W-interp car une détection parfaite ($D_{\mathrm{KL}} = +\infty$) est possible pour tout DWR en changeant $N_{\mathcal{S}}$ (comme pour DC-W-interp, cette limitation du cardinal de $\mathcal{S}$ peut se recalculer en réception).

Minimiser la distorsion à robustesse constante

A DNR et TEB fixés, une diminution de $N_{\mathcal{S}}$, Δ ou α permet de diminuer la distorsion. Plusieurs stratégies sont possibles : diminuer $N_{\mathcal{S}}$, chercher le couple (α, Δ) optimal à $N_{\mathcal{S}}$ fixé, ou bien effectuer une compensation des distorsions sans contrepartie sur Δ : si $m_l = 1$ on insère $x_k + \alpha(g(\underline{x}_k) - x_k)$, avec Δ fixe et α variable. Cette technique permet d'améliorer DWR de façon significative, au prix d'une grande perte

de performance de décodage (cf. *fig.* 4.10). On pourrait également combiner les trois techniques.

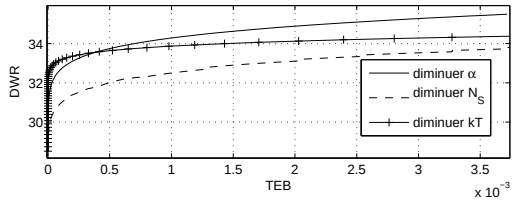

FIG. 4.10 – Maximisation de DWR en fonction du TEB, Lena, $L = 256$, WNR=-6 dB, $P_{\mathcal{S}} = 178$

Choix optimal de $(\Delta, N_{\mathcal{S}})$

De même, on peut calculer numériquement la valeur de Δ qui minimise le TEB à distorsion constante. Dans ce cas, $N_{\mathcal{S}} = \frac{2\sigma_x^2 N}{\Delta \sigma_U^2 \text{DWR}}$ donc il y a un compromis entre l'erreur d'interpolation et le nombre de points interpolés. Les *fig.* 4.11 à 4.13 montrent que les performances sont très peu améliorées par un choix optimal de $(\Delta, N_{\mathcal{S}})$. On se ramène donc à un compromis entre sécurité et imperceptibilité, la robustesse n'étant pas affectée. L'imperceptibilité est meilleure lorsque la technique d'interpolation n'est pas dénaturée ($\Delta = 0.25$), la sécurité est meilleure lorsque $N_{\mathcal{S}}$ diminue, ainsi que lorsque Δ augmente.

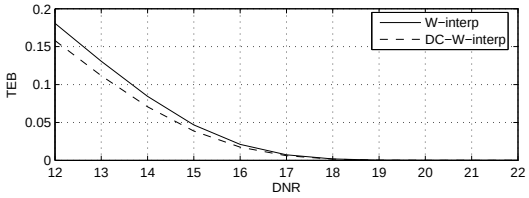

FIG. 4.11 – Amélioration par choix optimal de Δ, Lena, DWR=28 dB, $L = 256$

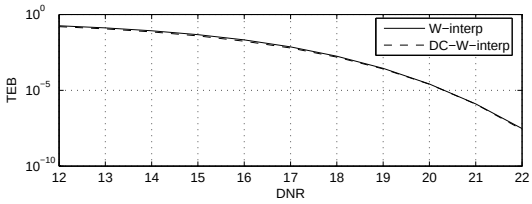

FIG. 4.12 – Amélioration par choix optimal de Δ, Lena, DWR=28 dB, $L = 256$

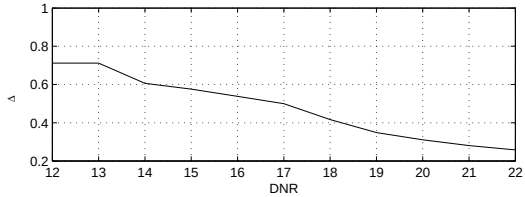

FIG. 4.13 – Choix optimal de Δ, Lena, DWR=28 dB, $L = 256$

4.4 Application à l'image : choix des paramètres et décodeur optimal

Dans cette partie, on étudie désormais l'application de W-interp à l'image. Deux implantations particulières de W-interp sont étudiées. On construit également un décodeur optimal prenant en compte la distribution de l'erreur d'interpolation en image.

4.4.1 Choix d'une grille d'interpolation

Dans le cadre général, $\mathcal{G}$ peut être quelconque. Une génération aléatoire de $\mathcal{G}$ pourrait d'ailleurs servir de paramètre de sécurité. Dans le cas d'une application à l'image, il est pourtant intéressant de structurer $\mathcal{G}$, bien que des techniques d'interpolation d'image existent sur des grilles dispersées. En effet, les techniques d'interpolation sur des grilles structurées ont de meilleures propriétés perceptuelles. De plus, le temps de calcul doit être raisonnable et il est intéressant d'effectuer une étude théorique des performances. On se limitera donc par la suite à des grilles équiréparties. Les grilles possibles restent très variées et dépendent du pas d'interpolation. Les points de $\mathcal{G}$ peuvent être équirépartis sur les lignes, les colonnes ou encore les diagonales.

Pour des raisons de symétrie (qui améliorent les performances et l'imperceptibilité), on choisira dans la suite la grille $\mathcal{G} = ((2\mathbb{Z}+1) \times 2\mathbb{Z}) \cup (2\mathbb{Z} \times (2\mathbb{Z}+1))$ qui a la forme d'un damier. On appelle parfois cette grille "grille en quinconce". Elle est également notée D_2 dans la notation de Conway&Sloane [MK05]. Les *fig.* 4.14 et 4.15 illustrent le fonctionnement de l'insertion de W-interp avec la grille en quinconce. La *fig.* 4.14 montre l'image originale (à gauche), les points de $\mathcal{G}$ correspondants (au centre), et le résultat d'une interpolation bilinéaire à partir de la grille (à droite). La *fig.* 4.15 montre une image originale (à gauche), le tatouage (au centre) et l'image tatouée (à droite). Les carrés bleus correspondent à l'insertion d'un bit -1 (pixels laissés identiques) et les carrés rouges à l'insertion d'un bit +1 (substitution par un pixel de l'image interpolée de la *fig.* 4.14).

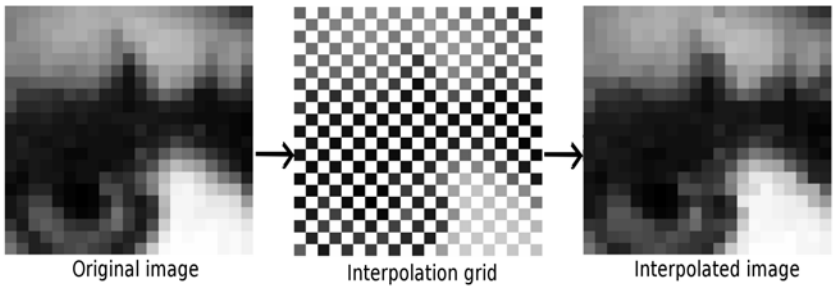

FIG. 4.14 – Interpolation dans W-interp : grille et image interpolée

4.4.2 Variante utilisant l'interpolation bilinéaire : W-bilin

Définition de W-bilin

Afin de donner une interprétation concrète aux paramètres de sécurité, on introduit une variante à la méthode de tatouage. Soit g une fonction d'interpolation "de base". Pour tout $(k_1, k_2) \in \mathcal{S}_l$ avec $m_l = +1$, on substitue par un point de coordonnées

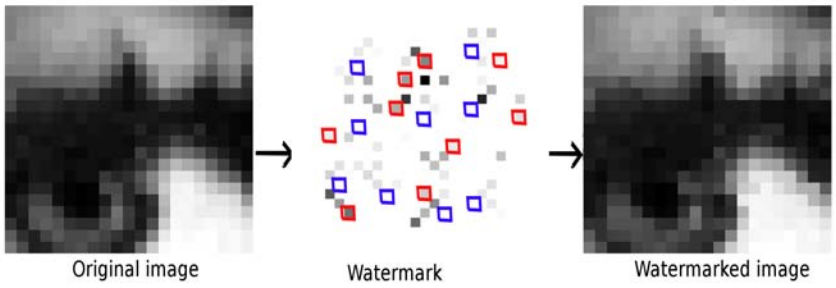

FIG. 4.15 – Tatouage par W-interp

décalées (cf. *fig.* 4.16)

$$g(\underline{x}_{k_1,k_2}) = \hat{x}(n_1 + \tau^u_{k_1,k_2}, n_2 + \tau^v_{k_1,k_2})$$

Les "décalages" $\tau^u_{k_1,k_2}$ et $\tau^v_{k_1,k_2}$ sont des variables aléatoires indépendantes.

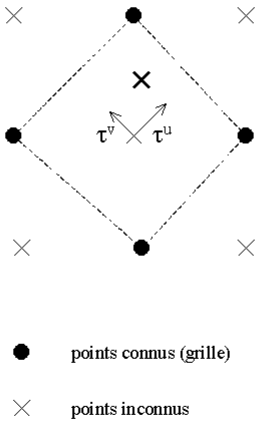

FIG. 4.16 – Décalages aléatoires des coordonnées

Calcul de Δ pour W-bilin

Ici, $N_v = 2$. Sans décalages, $\Delta = \frac{1}{4}$. Avec des décalages uniformément répartis sur $[-a, a]$, $g_{k,l}$ est du type $(\frac{1}{2} \pm \tau^u)(\frac{1}{2} \pm \tau^v)$. Donc

$$\Delta = \mathrm{E}[\left((\frac{1}{2}+\tau^u)(\frac{1}{2}-\tau^v)\right)^2 + \left((\frac{1}{2}+\tau^v)(\frac{1}{2}-\tau^u)\right)^2 + \left((\frac{1}{2}-\tau^u)(\frac{1}{2}-\tau^v)\right)^2 + \left((\tau^u+\frac{1}{2})(\frac{1}{2}+\tau^v)\right)^2]$$

$$\Delta = 4(\frac{1}{4} + \frac{a^2}{3})^2$$

Si $a = \frac{1}{2}$, $\Delta = 4/9 \simeq 0.444$.

4.4.3 Variante utilisant les splines : W-spline

Dans la seconde variante étudiée en image, g est dérivée de l'interpolation par B-splines cubiques. Les paramètres de sécurité sont identiques à ceux de W-bilin. Une différence importante avec W-bilin réside dans la taille du support de l'interpolant : 2x2 pour W-bilin, mais infini pour W-spline. On verra que cela a un impact important sur les performances théoriques et la robustesse.

Contrairement au masque du paragraphe 3.3.2, où tout le voisinage était connu, le support en damier se justifie ici : certains points du voisinage sont omis, (distinction entre $\mathcal{G}$ et $\mathcal{S}$). L'interpolation sur $\mathcal{G}$ est non séparable, mais une transformation lignes/diagonales rend possible l'implantation de l'interpolation par B-splines cubiques. En effet, les points connus de l'image sont équirépartis sur les diagonales (cf. *fig.* 4.17), donc en considérant l'image selon ses transformations selon les diagonales on peut appliquer l'interpolation (le filtrage devient séparable). De façon surprenante, on retrouve l'emploi d'une grille en quinconce et de l'interpolation bicubique [PF05] ou par B-spline [Sch00] dans le cadre des Matrices de Filtres Couleurs (*Color Filter Array*, CFA). En effet, la plupart des appareils photos utilisent des matrices de capteurs alternés. Chaque capteur correspond à une seule couleur (Rouge, Vert ou Bleu). Or si les capteurs bleus et rouges sont répartis sur une grille rectangulaire équirépartie, les capteurs verts sont répartis en quinconce (cf. *fig.* 4.18). Pour obtenir une image numérique à partir des capteurs, on effectue une interpolation. La solution proposée pour la couleur verte est identique à celle proposée ici [Sch00].

On s'attend à de meilleures propriétés perceptuelles pour W-spline que pour W-bilin. Cependant, l'introduction de $\mathcal{T}$ implique que W-bilin et W-spline ne réalisent pas une interpolation idéale, la différence perceptuelle est donc moindre qu'entre une interpolation linéaire classique et une interpolation par spline cubique.

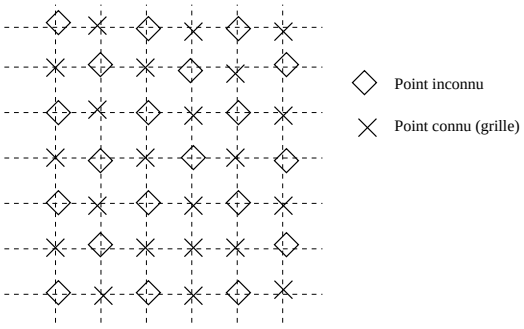

FIG. 4.17 – Grille d'interpolation

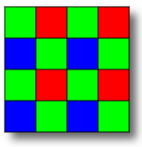

FIG. 4.18 – Exemple de matrice de filtres couleur CFA

Calcul de Δ pour W-spline

Pour W-spline, Δ est calculable théoriquement à partir de la spline cardinale d'ordre 3.

$$\Delta = \sum_{k=-\infty}^{+\infty} \sum_{l=-\infty}^{+\infty} \mathrm{E}[(\eta^3(\tau^u - k)\eta^3(\tau^v - l))^2]$$

Un calcul numérique donne $\Delta = 0.58$ pour W-spline sans décalages et $\Delta = 0.7665$ avec décalages. Ces constantes correspondent à celles observées expérimentalement.

W-spline sera donc moins robuste au AWGN que W-bilin. On aurait également $\Delta = 1$ pour une interpolation par plus proches voisins, et $\Delta \simeq 0.66$ pour la spline convolutive cubique. De manière générale, on pourrait penser que si $\sum_{j=1}^{N_v} g_j^k = 1$, $\Delta = \sum_{j=1}^{N_v} (g_j^k)^2$ devrait décroître avec N_v. Par exemple si $N_v = 4$ et que les g_j^k sont égaux, $\Delta = 1/16$. Cependant, les g_j^k ne sont pas nécessairement positifs. Notamment, pour les B-splines cubiques ou les spline convolutives cubiques, le noyau d'interpolation prend des valeurs négatives sur $]-2,-1[\cup]1,2[$. Donc il existe des coefficients supérieurs à 1/16 et Δ est grand.

4.4.4 Détecteur optimal sous l'hypothèse gaussienne généralisée

La partie 4.2.2 montre que le modèle gaussien permet de construire un détecteur simple et qui permet d'estimer fidèlement les performances expérimentales. Cependant, on a montré que dans le cas de l'application à une image, l'erreur d'interpolation suit une GGD (cf. partie 3.1.2). Le modèle GGD sera vérifié pour tout membre de la classe W-interp pour lequel g est une combinaison linéaire des pixels adjacents de somme des coefficients nulle. L'estimation du seuil théorique optimal peut donc être améliorée en suivant ce modèle, au prix de calculs plus complexes. En pratique, $\mathbf{y}$ est quantifiée en N_L niveaux de luminance à l'insertion ($N_L = 256$ par exemple pour une image de 8 bits). Le tatouage inséré en donc issu d'une version quantifiée de $\epsilon(\mathbf{x})$. Cependant, cette version quantifiée suit aussi une GGD (cf. *fig.* 4.19).

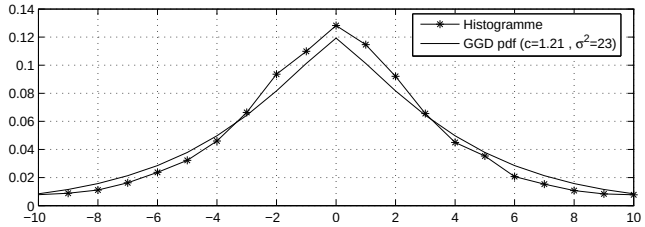

FIG. 4.19 – Histogramme de l'erreur d'interpolation quantifiée (Lena, 15000 pts, N_L=256)

Somme de deux gaussiennes généralisées

On désire modéliser par une gaussienne généralisée, la somme $\mathbf{R}$ de la variable aléatoire $\epsilon(\mathbf{x})$, de courbure $c_{\epsilon(\mathbf{x})}$ (proche de 1.5) et de variance $\sigma^2_{\epsilon(\mathbf{x})}$, avec le bruit gaussien $\mathbf{n}$ de courbure $c_{\epsilon(\mathbf{n})} = 2$ et de variance $\sigma^2_{\epsilon(\mathbf{n})}$. Comme les deux variables sont indépendantes, on a toujours $\sigma^2_R = \sigma^2_{\epsilon(\mathbf{n})} + \sigma^2_{\epsilon(\mathbf{x})}$. Le calcul de la courbure est plus compliqué. En effet, il n'existe pas d'expression théorique de la fonction caractéristique d'une GGD, même s'il est possible de la calculer numériquement [FKK04]. De plus, une somme de laplaciennes ($c = 1$) n'est pas forcément une laplacienne [Gre02]. On ne sait pas si une somme de GGD suit une GGD, ni calculer sa courbure si c'est le cas.

Ce problème peut être résolu numériquement. La densité de la somme est égale à la convolution des deux densités : $f_R = f_{\epsilon(\mathbf{x})} * f_{\epsilon(\mathbf{n})}$. Comme on connaît au décodage les paramètres de $\epsilon(\mathbf{x})$ et $\mathbf{n}$, on peut calculer numériquement f_R. On approche ensuite f_R par une GGD de paramètres c_R et σ_R^2 estimés par optimisation au sens des moindres carrés. Les Figs 4.20 et 4.21 montrent les bons résultats de la technique.

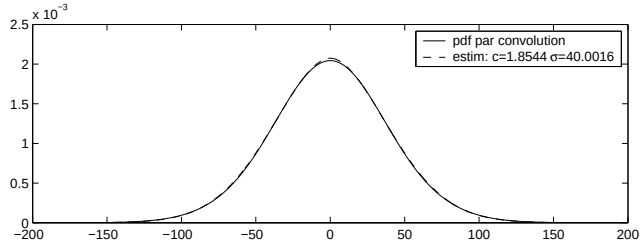

FIG. 4.20 – Pdf de la somme de deux GGD calculée par convolution et estimée par moindres carrés

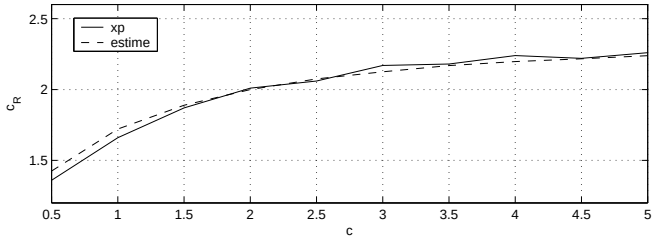

FIG. 4.21 – Courbure de la somme en fonction de $c_{\epsilon(\mathbf{x})}$: estimée et empirique, $\sigma^2_{\epsilon(\mathbf{n})} = 25$,$\sigma^2_{\epsilon(\mathbf{x})} = 25$

Détecteur optimal sous l'hypothèse gaussienne généralisée

Hypothèse H_1 (présence d'un tatouage) : $R \sim \mathcal{N}(0, (1+\Delta)\sigma^2_{\mathbf{n}})$, comme précédemment. Donc $R \sim \text{GGD}(0, c_1, \sigma_1^2)$ avec $c_1 = 2$ et $\sigma_1^2 = \sigma^2_{\epsilon(\mathbf{n})}$.

Hypothèse H_0 (absence de tatouage) :

$$r_k = \epsilon(x_k) + \epsilon(\mathbf{n})$$

$\epsilon(x_k)$ suit une loi gaussienne généralisée : $R \sim \text{GGD}(0, c_0, \sigma_0^2)$, avec $\sigma_0^2 = (1+\Delta)\sigma^2_{\mathbf{n}} + \sigma^2_{\epsilon(\mathbf{x})}$ et c_0 estimée comme précédemment.

Statistique de test T : on peut ré-exprimer ces deux lois en fonction des paramètres A, β, c. La statistique de test de Neyman-Pearson est alors :

$$\ln\left(\frac{\prod_k A_1 e^{-|\alpha_1 r_k|^{c_1}}}{\prod_k A_0 e^{-|\alpha_1 r_k|^{c_0}}}\right) = P_{\mathcal{S}} \ln\left(\frac{A_1}{A_0}\right) - \sum_{i=1}^{P_{\mathcal{S}}} (|\beta_1 r_i|^{c_1} - |\beta_0 r_i|^{c_0})$$

ce qui conduit à

$$\mathbf{T} = \sum_{i=1}^{P_{\mathcal{S}}} (|\alpha_1 r_i|^{c_1} - |\alpha_0 r_i|^{c_0}) \tag{4.3}$$

Si $c_0 = c_1 = 2$, on retrouve la statistique de test calculée pour le cas gaussien.

On approche ensuite $\mathbf{T}$ par une loi normale dont on estime les paramètres $\mu_{\mathbf{T}}$ et $\sigma^2_{\mathbf{T}}$. Si le TEB est très faible, cette approximation justifiée par le théorème Central-Limite peut conduire à une estimation peu précise.

Espérance de T : calculons tout d'abord E[$(|R|^{c_a})|H_b$] avec $(a,b) \in \{0,1\}^2$ (sous l'hypothèse H_b, $R \sim$ GGD$(0, c_b, A_b, \alpha_b)$).
Soit $\Psi : x \to y = |x|^{c_a}$. Alors $f_Y(y) = |(\Psi^{-1})'(y)| f_X(\Psi^{-1}(y))$,
où sur $[0, +\infty[$: $\Psi^{-1}(y) = y^{1/c_a}$ et $(\Psi^{-1})'(y) = \frac{1}{c_a} y^{1/c_a - 1}$.

$$\begin{aligned}
\int_0^{+\infty} y f_{(|R|^{c_a})|H_b}(y)\mathrm{d}y &= \int_0^{+\infty} y \frac{A_b}{c_a} e^{-|\alpha_b|^{c_b} y^{c_b/c_a}} y^{1/c_a - 1}\mathrm{d}y \\
&=_{u=y^{c_b/c_a}} \int_0^{+\infty} u^{c_a/c_b} \frac{A_b}{c_a} e^{-|\alpha_b|^{c_b} u} u^{(c_a/c_b)(1/c_a - 1)} \frac{c_a}{c_b} u^{c_a/c_b - 1}\mathrm{d}u \\
&=_{t=u|\alpha_b|^{c_b}} \int_0^{+\infty} \frac{t^{c_a/c_b}}{|\alpha_b|^{c_a}} \frac{A_b}{c_b} \frac{1}{|\alpha_b|^{c_b}} e^{-t} \frac{1}{(|\alpha_b|^{c_b})^{1/c_b - 1}} t^{1/c_b - 1}\mathrm{d}t \\
&= \frac{A_b}{c_b |\alpha_b|^{1+c_a}} \Gamma\left(\frac{1+c_a}{c_b}\right)
\end{aligned}$$

avec $\Gamma(x) = \int_0^{+\infty} e^{-t} t^{x-1}\mathrm{d}t$. Soit $\Lambda(A_b, \alpha_b, c_b, c_a) = 2\frac{A_b}{c_b|\alpha_b|^{1+c_a}}\Gamma(\frac{1+c_a}{c_b})$, et soit pour simplifier $\Lambda^*(b, c_a) = \Lambda(A_b, \alpha_b, c_b, c_a)$. Donc E[$(|R|^{c_a})|H_b$]=$\Lambda^*(b, c_a)$.
Finalement,

$$\mu_{\mathbf{T}|H_b} = \mathrm{E}[\mathbf{T}|H_b] P_{\mathcal{S}} (|\alpha_1|^{c_1} \Lambda^*(b, c_1) - |\alpha_0|^{c_0} \Lambda^*(b, c_0))$$

Variance de T : après développement de

$$\mathbf{T}^2 = \sum_{i=1}^{P_{\mathcal{S}}} \sum_{j=1}^{P_{\mathcal{S}}} (|\alpha_1 r_i|^{c_1} - |\alpha_0 r_i|^{c_0})(|\alpha_1 r_j|^{c_1} - |\alpha_0 r_j|^{c_0})$$

on a :

$$\begin{aligned}
\sigma^2_{\mathbf{T}|H_b} =& P_{\mathcal{S}}(P_{\mathcal{S}} - 1) \left(|\alpha_1|^{2c_1} \Lambda^*(b, c_1)^2 + |\alpha_0|^{2c_0} \Lambda^*(b, c_0)^2 - 2|\alpha_1|^{c_1} |\alpha_0|^{c_0} \Lambda^*(b, c_1) \Lambda^*(b, c_0)\right) \\
&+ P_{\mathcal{S}} \left(|\alpha_1|^{2c_1} \Lambda^*(b, 2c_1) + |\alpha_0|^{2c_0} \Lambda^*(b, 2c_0) - 2|\alpha_1|^{c_1} |\alpha_0|^{c_0} \Lambda^*(b, c_1 + c_0)\right) - \mu^2_{\mathbf{T}|H_b}
\end{aligned}$$

Seuil de décision optimal : soit $Q(x) = \int_x^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$ et $Q^{-1}(x)$ sa fonction réciproque. Comme auparavant, à P_{fa} donnée, $\eta = \mathrm{F}^{-1}_{\mathbf{T}|H_1}(1 - \mathrm{P}_{\text{fa}}) = \sigma_{\mathbf{T}|H_1} Q^{-1}(\mathrm{P}_{\text{fa}}) + \mu_{\mathbf{T}|H_1}$ pour la détection et $P_{\text{nd}} = \mathrm{F}_{T|H_0}(\eta) = 1 - Q(\frac{\eta - \mu_{\mathbf{T}|H_0}}{\sigma_{\mathbf{T}|H_0}})$.
De plus, P_{fa} correspond à $P_{\text{fa}} = 1 - \mathrm{F}_{\mathbf{T}|H_0}(\eta) = Q(\frac{\eta - \mu_{\mathbf{T}|H_1}}{\sigma_{\mathbf{T}|H_1}})$.

Le seuil optimal pour le décodage se calcule par

$$\eta_{\text{th}} \text{ tel que } \frac{\partial \text{TEB}}{\partial \eta}(\eta_{\text{th}}) = 0$$

soit

$$\eta_{\text{th}} \text{ tel que } f_{T|H_0}(\eta_{\text{th}}) = f_{T|H_1}(\eta_{\text{th}})$$

Donc

$$\eta_{\text{th}} = \frac{\sigma_{\mathbf{T}|H_1}\mu_{\mathbf{T}|H_0} + \sigma_{\mathbf{T}|H_0}\mu_{\mathbf{T}|H_1}}{\sigma_{\mathbf{T}|H_0} + \sigma_{\mathbf{T}|H_1}}$$

Performances à la détection : on a vu qu'il était possible de calculer une borne théorique de capacité de décodage. La distance de Kullback-Leibler (cf. paragraphe 1.5.3) peut quant à elle jouer le rôle de "capacité théorique de détection" [DFHS03]. En effet, la distance de Kullback-Leibler permet de mesurer la distance entre distributions. Par exemple, si $P_{\text{fa}} = 0$, $P_d \leq 1 - e^{-D_{\text{KL}}}$ [DFHS03]. Nous reproduisons ici la comparaison des distances de Kullback-Leibler D_{KL} de [DFHS03], ainsi que l'espace de détection associé. On y a rajouté les performances de LISS en fonction de λ.

Méthode	D_{KL} sans bruit	D_{KL} avec bruit
DS	$\frac{1}{2}N\frac{1}{\text{DWR}}$	$\frac{1}{2}N\frac{\sigma^2_{\mathbf{w}}}{\sigma^2_{\mathbf{x}}+\sigma^2_{\mathbf{n}}}$
JANIS	$\frac{1}{2}n_oN\frac{1}{\text{DWR}}$	$\frac{1}{2}\left(\log\frac{\sigma^2_{T\|H_1}}{\sigma^2_{T\|H_0}} - 1 + \frac{\sigma^2_{T\|H0}}{\sigma^2_{T\|H_1}} + \frac{\mu^2_{T\|H_1}}{\sigma^2_{T\|H_1}}\right)$
ZATT	$+\infty$	$\frac{1}{2}N\frac{1}{\text{DWR}}\left(\log(\frac{\sigma^2_{\mathbf{n}}}{\sigma^2_{\mathbf{n}}+\sigma^2_{\mathbf{x}}}) - 1 + \frac{\sigma^2_{\mathbf{n}}+\sigma^2_{\mathbf{x}}}{\sigma^2_{\mathbf{n}}}\right)$
PEAK	$+\infty$	$\frac{1}{2}\left(\log(\frac{\sigma^2_{\mathbf{n}}}{\sigma^2_{\mathbf{n}}+\sigma^2_{\mathbf{x}}}) - 1 + \frac{\sigma^2_{\mathbf{n}}+\sigma^2_{\mathbf{x}}}{\sigma^2_{\mathbf{n}}} + \frac{N\sigma^2_{\mathbf{w}}-\sigma^2_{\mathbf{x}}}{\sigma^2_{\mathbf{n}}}\right)$
LISS	$+\infty$	$\frac{1}{2}\left(\log(\frac{\sigma^2_{\mathbf{n}}+(1-\lambda^2)\sigma^2_{\mathbf{x}}}{\sigma^2_{\mathbf{n}}+\sigma^2_{\mathbf{x}}}) - 1 + \frac{\sigma^2_{\mathbf{n}}+\sigma^2_{\mathbf{x}}}{\sigma^2_{\mathbf{n}}+(1-\lambda^2)\sigma^2_{\mathbf{x}}} + \frac{N\sigma^2_{\mathbf{w}}-\lambda^2\sigma^2_{\mathbf{x}}}{\sigma^2_{\mathbf{n}}+(1-\lambda^2)\sigma^2_{\mathbf{x}}}\right)$

Pour W-interp, l'"espace de tatouage" est l'erreur d'interpolation $\epsilon(\mathbf{x})$ et l'espace de détection est la puissance de $\epsilon(\mathbf{y})$. En l'absence de bruit, on compare une distribution continue au singleton 0 et $D_{\text{KL}} = +\infty$ [DFHS03]. En présence de bruit, on compare $\mathcal{N}(\mu_{\mathbf{T}|H_0}, \sigma^2_{\mathbf{T}|H_0})$ et $\mathcal{N}(\mu_{\mathbf{T}|H_1}, \sigma^2_{\mathbf{T}|H_1})$:

Méthode	espace	D_{KL} sans bruit	D_{KL} avec bruit
W-interp	$\mathbb{R}$	$+\infty$	$\frac{1}{2}\left(\log\frac{\sigma^2_{\mathbf{T}\|H_1}}{\sigma^2_{\mathbf{T}\|H_0}} + \frac{1}{\sigma^2_{\mathbf{T}\|H_1}}((\mu_{\mathbf{T}\|H_0} - \mu_{\mathbf{T}\|H_1})^2 + \sigma^2_{\mathbf{T}\|H_0} - \sigma^2_{\mathbf{T}\|H_1})\right)$

La *fig.* 4.22 montre une comparaison de D_{KL} pour les différentes techniques. Les techniques DS et DS+W ont une performance bornée par les interférences de l'image hôte. Pour les techniques de tatouage informé W-interp et LISS, $D_{\text{KL}} \to +\infty$ lorsque le bruit diminue. W-interp est la meilleure technique en cas de bruit faible. Par contre, W-interp n'est pas robuste à un fort bruit (elle est même dépassée par DS). Lorsque le facteur d'étalement est plus faible, on remarque à nouveau que LISS n'offre plus de bonnes performances, et que l'utilisation d'un filtrage de Wiener est nécessaire.

Performances au décodage : Le détecteur peut être facilement étendu au décodage comme pour le détecteur gaussien. La statistique de test $\mathbf{T} = \sum_{i=1}^{P_S}(|\beta_1 r_i|^{c_1} - |\beta_0 r_i|^{c_0})$ n'est utilisable que dans le scenario de bruit AWGN connu au décodage car elle utilise c_0. Le seuil empirique utilisé dans les autres scénarios est donc dérivé de la statistique de test du cas gaussien.

Les simulations montrent que les performances théoriques du décodeur optimal correspondent exactement aux performances pratiques. De plus, une légère erreur sur la courbure c_0 entraîne une erreur importante au décodage, ce qui confirme à la fois la validité du modèle et celle des calculs.

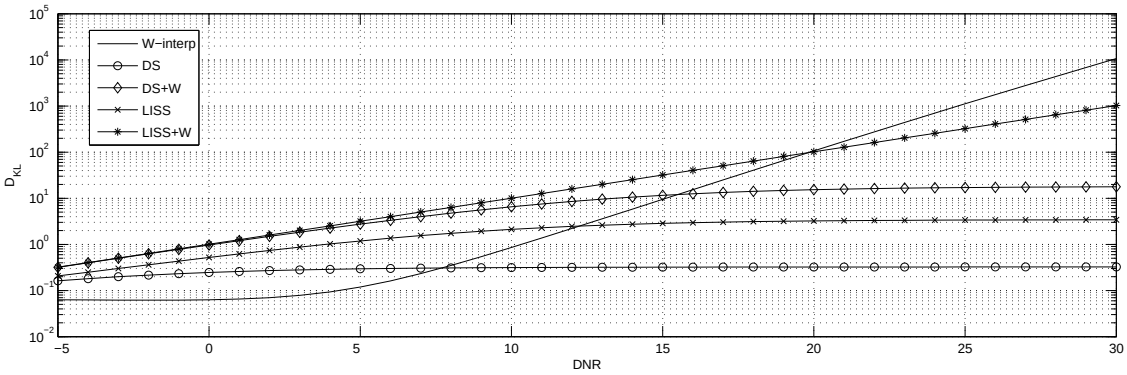

FIG. 4.22 – Comparaison des D_{KL}, Lena, DWR=28 dB, $N = 2^{12}$

La comparaison avec le décodeur classique (cf. *fig.* 4.23 et 4.24) montre des performances proches, quels que soient les paramètres (L, P, WNR). En effet, le décodeur gaussien, bien que non-optimal, permet une adéquation avec les performances expérimentales dans la plupart des configurations, ce qui souligne la robustesse statistique de la technique. Plus le bruit est important, plus les deux décodeurs sont proches ($f_{R|H_0}$ tend vers une gaussienne). Lorsque P_S est très petit, les deux décodeurs restent efficaces même si les approximations de T suivant le χ^P_2, dans un cas, et une gaussienne dans l'autre, ne sont plus valables. Sur les images Poivrons et Pentagone, qui présentent une courbure c plus faible (et donc plus éloignée du cas gaussien), on constate encore mieux la différence entre les décodeurs (le décodeur optimal modélise mieux la distribution $P_e(\eta)$). En effet, la moyenne et la variance de la loi du χ^P_2 sont liées et dépendent de P et de la variance. Pour le décodeur optimal, on a utilisé une approximation gaussienne de $\mathbf{T}$ (équation (4.3)), dont la moyenne et la variance dépendent en réalité de c, P et σ^2 de façon plus complexe. On peut donc modéliser la statistique de test plus précisément. Par contre, le seuil optimal est très peu affecté.

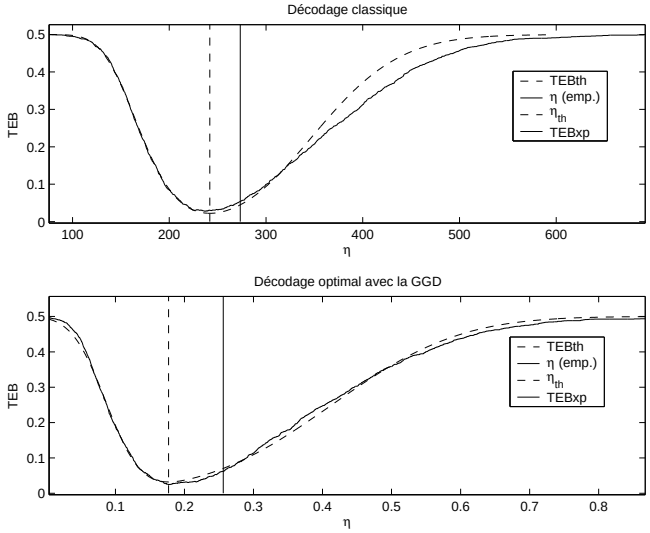

FIG. 4.23 – Comparaison entre décodeur classique et optimal, Bateaux, $L = 2048$, DWR=33, WNR= -2

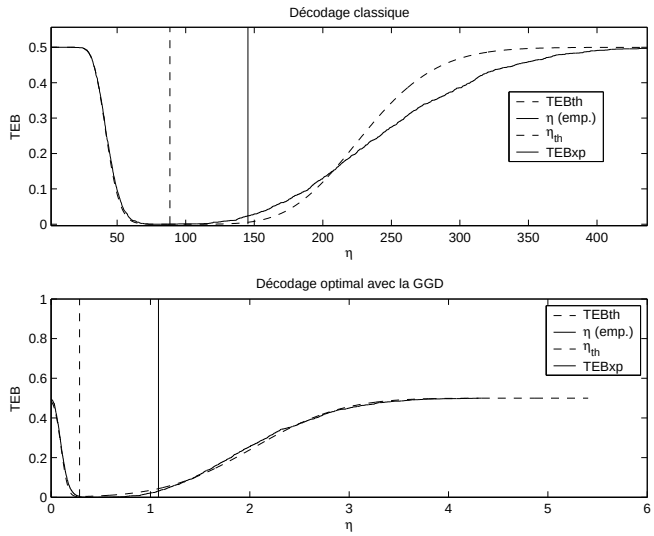

FIG. 4.24 – Comparaison entre décodeur classique et optimal, Bateaux, $L = 2048$, DWR=33, WNR= -2

4.5 Application à l'image : étude de l'imperceptibilité

4.5.1 Etude du tatouage

W-interp utilisant des propriétés intrinsèques à l'image $\mathbf{x}$ telles que la corrélation entre pixels voisins, son efficacité dépend elle aussi de $\mathbf{x}$. Ainsi, le DWR minimum que W-interp peut atteindre est $\frac{4\sigma^2_{\mathbf{x}}}{\sigma^2_{\epsilon(\mathbf{x})}}$. Les propriétés de robustesse et de sécurité de W-interp sont donc limitées sur certaines images. Des images spécifiques pourraient également conduire à des tatouages perceptibles ou à des artefacts d'interpolation. Cependant, sur un ensemble de 50 images naturelles, on observe que $\sigma^2_{\epsilon(\mathbf{x})} > 20$. En moyenne, $\sigma^2_{\epsilon(\mathbf{x})} = 108$ et $c_{\epsilon(\mathbf{x})} = 0.6$. Alors $\sigma^2_{\mathbf{x}}/\sigma^2_{\epsilon(\mathbf{x})} = 15.4$ dB en moyenne. De plus, la proportion de points où l'erreur d'interpolation (donc le tatouage) est nulle est seulement de 8% en moyenne.

Les *fig.* 4.25 à 4.28 montrent des exemples de tatouages générés par W-interp.

FIG. 4.25 – Image Lena originale, tatouée et tatouage (détails), W-bilin, DWR=20

FIG. 4.26 – Image Lena originale, tatouée et tatouage (détails), W-bilin, DWR=28

Apparition d'artefacts : Lorsque le signal ne suit pas le modèle adéquat, l'interpolation conduit classiquement à des artefacts (oscillations, repliement, effet de bloc,

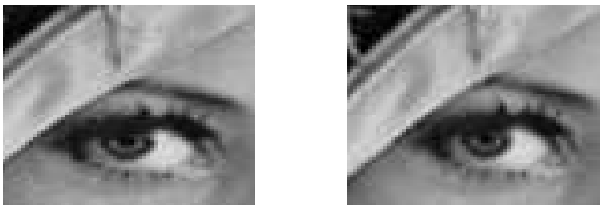

FIG. 4.27 – Lena : image originale et image tatouée (détails), W-bilin

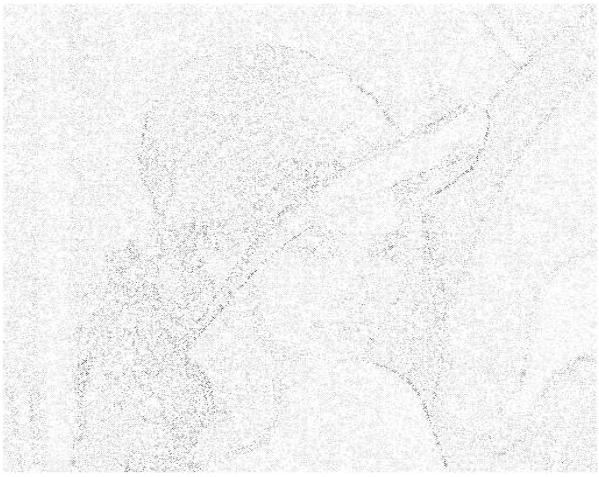

FIG. 4.28 – Lena : tatouage, W-bilin

effet de flou) [TBU00]. Empiriquement, nous n'avons cependant pas constaté d'apparition de ces artefacts par tatouage avec W-bilin et W-spline. Ceci est dû à l'utilisation d'images naturelles, d'une puissance d'insertion raisonnable, de décalages aléatoires et au fait de ne substituer qu'une partie des points. Notamment, l'utilisation d'une grille en quinconce pourrait conduire à un effet de flou sur les contours d'une image, similaire à celui introduit par une séquence sous-échantillonnage/suréchantillonnage. Contrairement au cas des masques proposés dans la partie 3.3, W-interp pourrait introduire de tels artefacts car on y insère le résultat exact d'une interpolation. Cependant, dans W-interp, seule une partie des points est modifiée à l'insertion (par exemple 15% des points à 28 dB). De plus, la fonction interpolante est différente pour chaque pixel interpolé. Grâce à ces deux propriétés, nous n'avons pas constaté en pratique d'apparition d'artefacts au niveau des contours.

Une étude subjective telle que celle proposée en annexe C.1 met en évidence des effets de pixellisation près des contours de l'image tatouée lorsqu'on remplace certains points de l'image par le résultat d'une interpolation, pour les techniques d'interpolation les plus simples. W-spline semble moins vulnérable à ce phénomène. Afin de se mettre totalement à l'abri de ces artefacts, on devrait combiner W-interp à un masque perceptuel fondé sur une détection des contours [CP95], de façon à concentrer les modifications sur les textures. Il serait également possible de limiter l'amplitude maximale des distorsions en excluant les points où l'erreur d'interpolation est trop grande. Une attaque pourrait cependant modifier ce seuil de perceptibilité, ce qui pourrait empêcher W-interp d'atteindre des TEB très faibles. Cependant, sans modifier le principe de W-interp, il est également possible de choisir comme fonction g une technique d'interpolation adaptative préservant les contours (cf. partie 3.1).

D'éventuels artefacts échappent souvent aux mesures perceptuelles objectives uti-

lisées dans la suite. De plus, il est difficile de valider une étude subjective. Nous avons donc fait le choix de ne pas implanter cette variante de W-interp dans ce rapport, tout en encourageant la combinaison de W-interp avec des techniques d'interpolation plus complexes.

Modification de l'histogramme : dans le cas d'images quantifiées grossièrement, l'interpolation est réputée modifier l'histogramme de l'image en créant des niveaux de gris lorsqu'ils étaient absents, ce qui pourrait permettre de détecter la présence d'un tatouage. Cependant, sur les images "naturelles" étudiées, les niveaux de gris sont déjà tous occupés au sein de la dynamique de l'image. W-interp n'est donc pas particulièrement vulnérable de ce point de vue.

Localisation des déformations : W-interp agit sur un grand nombre de points. La puissance globale du tatouage est donc répartie sur l'image, comme pour la méthode DS. Ainsi, pour Lena et DWR=28 dB, 17% des points de l'image sont modifiés. Cependant, la technique DS doit utiliser un masque psychovisuel (cf. paragraphe 1.5.3) pour garantir l'imperceptibilité, en plus de cette répartition de puissance. Dans le cas de W-interp, on observe empiriquement que les points où l'erreur est la plus grande correspondent aux contours ou à une grande variance locale de l'image. Ceci est logique car on modifie les composantes passe-haut de l'image. Pour Lena, W-bilin et DWR=28 dB, les 15000 points (*i.e.* $3,8\%$) les plus modifiés se situent sur les contours.

On sait que ces zones sont moins sensibles perceptuellement à une déformation importante. Par exemple, ce sont celles que l'on modifie en priorité en utilisant le masque NVF (cf. paragraphe 1.5.3). Par exemple, pour Lena, W-bilin et DWR=28 dB, NVF et l'interpolation ont 40% de leurs 5000 points les plus modifiés en commun. Aux points où la modification par W-interp est plus faible, on observe que $-3 < \epsilon(\mathbf{x}) < 3$, ce qui correspondrait à un DWR de 23 dB si tous les points étaient tatoués. Pour les 94% des points les moins modifiés, $-2,5 < \epsilon(\mathbf{x}) < 2,5$, soit DWR=24,5 au pire. Globalement, on peut donc dire que le masquage perceptuel intrinsèque à l'interpolation répond au même critère d'imperceptibilité que le masque NVF, même si son principe est différent.

FIG. 4.29 – Points modifiés par W-interp et par le masque NVF (10000 plus grandes valeurs)

Spectre du tatouage

Le spectre du tatouage a la forme d'un "lobe" centré sur les moyennes fréquences (cf. *fig.* 4.30). Il s'agit d'un filtrage passe-haut du spectre de l'image. La forme de lobe du spectre vient de la multiplication du spectre de l'image (en forme de pic) par un noyau d'erreur d'interpolation qui correspond à un filtrage passe-haut. Plus l'ordre de la technique d'interpolation est élevé, moins il y a d'oscillations et plus la pente est élevée [TBU00]. Si l'ordre $\rightarrow +\infty$, la réponse impulsionnelle est un sinus cardinal et le spectre de l'image est multiplié par une fenêtre passe-haut idéale.

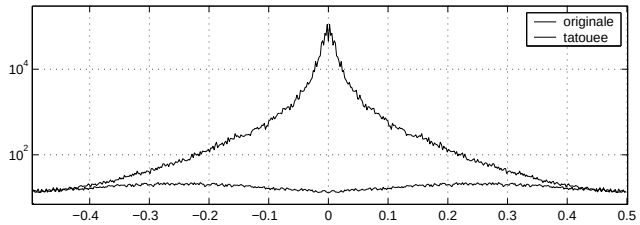

FIG. 4.30 – Spectres de l'image originale (en haut) et du tatouage (en bas), W-bilin, Lena, DWR=22 dB

Intercorrélation entre le tatouage et l'image

Le tatouage généré par W-interp est très corrélé à l'image. La *fig.* 4.31 montre les cartes d'intercorrélation 2D entre $\mathbf{w}$ et $\mathbf{x}$ de plusieurs techniques de tatouage. Une forte valeur est représentée par un point sombre. La puissance de l'intercorrélation pour W-interp est concentrée sur le point central (à peine visible). Cette propriété sera utile lors de l'étude de la robustesse au débruitage (cf. paragraphe 4.6).

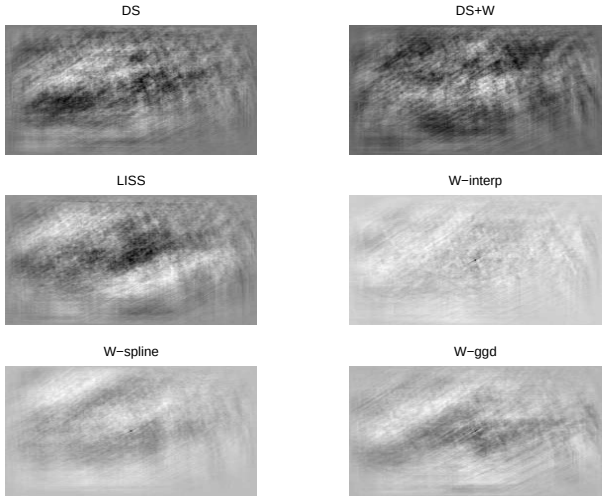

FIG. 4.31 – Comparaison des cartes d'intercorrélation 2D entre $\mathbf{w}$ et $\mathbf{x}$

Justification objective de l'imperceptibilité

La méthode W-interp est construite sur l'hypothèse que l'interpolation déforme peu l'image visuellement. Dans cette section, on vérifie que cette condition est bien respectée. L'imperceptibilité est évaluée de manière objective par une mesure de qualité perceptuelle (cf. paragraphe 1.5.3). W-interp garantit intrinsèquement le respect du critère de PSNR car le nombre N_S de points modifiés est choisi en fonction du DWR. Il est donc aisé de respecter le critère PSNR<36 dB. W-interp ne respecte pas strictement le critère des JND (cf. paragraphe 1.5.3). Cependant, de nombreux auteurs ont déjà indiqué que cette contrainte est trop restrictive vis-à-vis de la puissance d'insertion.

La distance de Watson D_W varie selon l'itération, on présente donc ici les valeurs moyennes. Si l'on normalise la distance de Watson par le nombre de pixels N, on obtient en moyenne 0.0015 (DS) et 4.10^{-5} (DS+DCT). D_W diminue environ de moitié quand DWR diminue d'environ 6.5 dB. Par exemple, pour Lena, $D_W = 390$ si DWR=28 dB, $D_W = 210$ si DWR=34.5 dB, $D_W = 91$ si DWR=41 dB, $D_W = 45$ si DWR=47.5 dB. Les performances de W-interp sont décevantes selon cette mesure

objective. Cependant, la distance de Watson est extrêmement biaisée à l'égard du masque perceptuel associé. Elle ne reflète pas l'amélioration perceptuelle apportée par les masques spatiaux, quels qu'ils soient.

La distance SSIM est comprise entre 0 et 1, 1 étant le maximum de fidélité perceptuelle. Les résultats sont très favorables à W-interp : W-interp obtient une bien meilleure qualité perceptuelle que DS avec les masques NVF et DCT (Ahumada et al.). Seul le masque d'Alvarez est légèrement meilleur. Selon ce critère, l'utilisation de W-interp correspond à un gain d'environ 4 dB en DWR, à distance SSIM constante (contre 2,5 dB pour le masque NVF).

La table ci-dessous présente les résultats pour PSNR=43,5 dB. Selon les mesures objectives, W-spline apporte peu d'amélioration perceptuelle par rapport à W-bilin, contrairement à ce qu'on aurait pu attendre. Ceci est dû à l'utilisation des décalages aléatoires qui uniformisent les performances, mais également au fait que la grille d'interpolation $\mathcal{G}$ omet certains points du voisinage, si bien que l'interpolation est moins bonne que dans l'utilisation classique des splines. Ce résultat mériterait d'être complété par des études subjectives.

	SSIM	D_W	D_{KL}
DS	0.9870	365	1370
DS+NVF	0.9944	317	709
DS+masque DCT	0.9953	6	295
W-bilin $a = 0$	0.9981	322	9.26
W-spline $a = 0$	0.9989	295	8.17
W-bilin $a = 1/2$		330	
W-spline $a = 1/2$		321	
W-bilin $a = 1$	0.9968	357	9.54
W-spline $a = 1$	0.9968	354	8.78

4.6 Application à l'image : étude de la robustesse

Dans cette section, on compare expérimentalement W-bilin et W-spline à DS+W et DS, ST-SCS et LISS, ainsi qu'à SCS et RDM, qui utilisent aussi une redondance par répétition. On étudie la robustesse à diverses attaques dans le scénario où l'attaque n'est pas connue à l'insertion. On utilise donc le seuil empirique ou itératif dans W-interp (cf. annexe C.2). Il n'y a pas de compensation des distorsions dans SCS, ST-SCS ($\alpha = 1$) et LISS ($\lambda = 1$). Cependant, pour les valeurs de DWR, WNR et L utilisées, α dans ST-SCS et λ dans LISS auraient été proches de 1 (cf. paragraphes 1.3.2 et 1.3.1), ces deux méthodes ne sont donc pas défavorisées dans la comparaison. Dans le scénario où l'attaque est connue du décodeur, une étude théorique du seuil de décision et des performances serait possible pour W-interp et certaines attaques. Elle conduirait à un choix du seuil plus précis, donc à de meilleures performances de W-interp.

4.6.1 Tatouage haut débit

La *fig.* 4.32 montre les performances de W-interp en fonction de la charge utile. Pour $P > 64$, ce qui correspond à $L < 4096$ si $N = 2^{18}$, BER$< 10^{-5}$. W-interp rejette donc presque les interférences de l'hôte dans un scénario de tatouage bas débit. Les performances de W-interp reposent sur l'utilisation de la redondance dans la règle de décodage. W-interp est donc peu performant dans un scénario de tatouage haut débit.

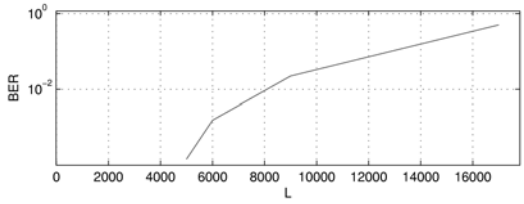

FIG. 4.32 – Influence de la redondance sur le décodage, $\sigma^2_{\mathbf{n}} = 0$

4.6.2 Bruit AWGN

Le scénario choisi dans cette section est celui où $\sigma^2_{\mathbf{n}}$ est connue à l'insertion. On peut donc utiliser le seuil théorique η_{th} dans W-interp, et effectuer une compensation des distorsions dans LISS et ST-SCS. Même si l'erreur d'interpolation n'est pas exactement gaussienne, les *fig.* 4.33 et 4.34 montrent que les résultats expérimentaux sont très proches de la courbe théorique.

W-interp est à rejet des interférences de l'hôte, donc à $\sigma^2_{\mathbf{n}}$ faible, les performances sont bien meilleures que celles de DS et DS+W (cf. *fig.* 4.34). Si $\sigma^2_{\mathbf{n}}$ est très grand, DS est meilleur, ce qui est logique car cette méthode est conçue pour résister au mieux à bruit additif gaussien indépendant de l'image. On vérifie expérimentalement que W-spline est moins robuste que W-bilin à l'attaque AWGN (cf. *fig.* 4.34 et 4.33). On notera les excellentes performances de SCS, RDM et ST-SCS (cf. partie 1.3.2) construits pour résister à l'attaque AWGN, qui n'est pas affectée par l'image hôte, et qui prennent en compte $\sigma^2_{\mathbf{n}}$ à l'insertion. Le cas de LISS est particulier. Cette méthode rejette les interférences de l'hôte si $P > \frac{\sigma^2_{\mathbf{x}}}{\sigma^2_{\mathbf{w}}}$ (cf. partie 1.3.1). Selon DWR, si L est faible LISS est robuste au bruit AWGN (cf. *fig.* 4.33), mais si L est faible W-interp offre de meilleures performances. On observe cette propriété autour de $L = 300$ si DWR=28 dB. La méthode LISS+W fournit des performances encore meilleures.

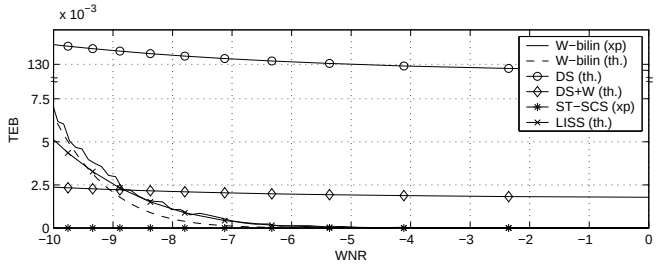

FIG. 4.33 – Robustesse à l'ajout de bruit gaussien en fonction de WNR, DWR=28, $L = 300$

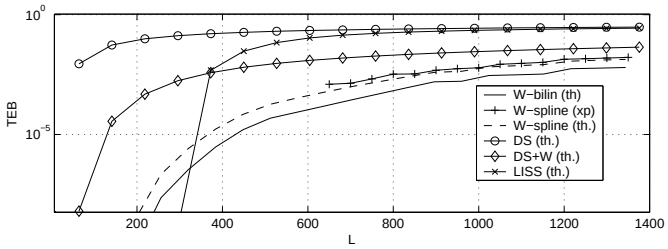

FIG. 4.34 – Robustesse à l'ajout de bruit gaussien en fonction de L, WNR=-4 dB, DWR=28 dB

4.6.3 Attaques classiques

Cette section s'intéresse aux attaques de type "traitement du signal" classiques : compression JPEG, filtrage de Wiener, égalisation d'histogramme. La *fig.* 4.35 montre l'effet d'une attaque sur les distribution des erreurs d'interpolation dans les deux hypothèses. La source d'erreur de décodage provient du rapprochement des distributions des deux classes.

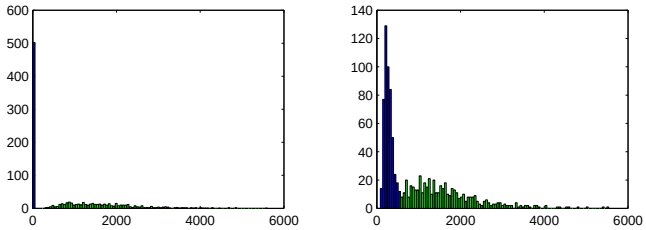

FIG. 4.35 – Histogramme de ρ_l^2, faible compression JPEG : (gauche) avant attaque, (droite) après attaque

Compression JPEG

La robustesse de W-interp à la compression JPEG est bonne (cf. *fig.* 4.36) : seul DS+W obtient de meilleures performances, et W-interp n'est affectée que pour un facteur de qualité faible ($Q < 85\%$). RDM et SCS, non représentés ici, sont les moins robustes à la compression JPEG. W-bilin est plus robuste que W-spline à la compression JPEG. La plus grande robustesse de W-interp à la compression JPEG vient du fait que cette méthode privilégie plus les moyennes fréquences (cf. paragraphe 4.5.1), qui sont moins affectées que les hautes fréquences par la quantification. On pourrait s'inspirer de [EG01][FKK04] pour calculer les performances théoriques de W-interp face à la compression JPEG (ou à une quantification simple).

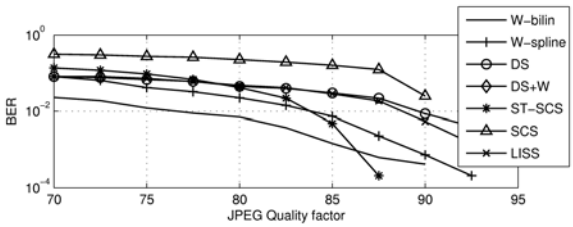

FIG. 4.36 – Robustesse de W-interp à la compression JPEG, $L = 64$, DWR=28 dB

Débruitage

La robustesse de W-interp au débruitage par filtrage de Wiener est bien meilleure que celle des méthodes DS (cf. *fig.* 4.37). Cela est dû au fait que le débruitage tente d'éliminer le tatouage en le considérant comme un bruit additif gaussien, indépendant de l'image. Cette hypothèse est vraie pour les méthodes DS et DS+W. Pour LISS et ST-SCS, le tatouage n'est pas totalement indépendant de l'image mais ils restent peu corrélés (cf. paragraphe 4.5.1). Pour W-interp, l'hypothèse n'est pas vérifiée : le tatouage n'est pas gaussien (c'est un mélange de valeurs nulles et d'une variable qui suit une GGD), et surtout il est fortement corrélé à l'image. De plus, le spectre du tatouage n'est pas plat (cf. paragraphe 4.5.1). Les performances de W-bilin et W-spline sont ici proches.

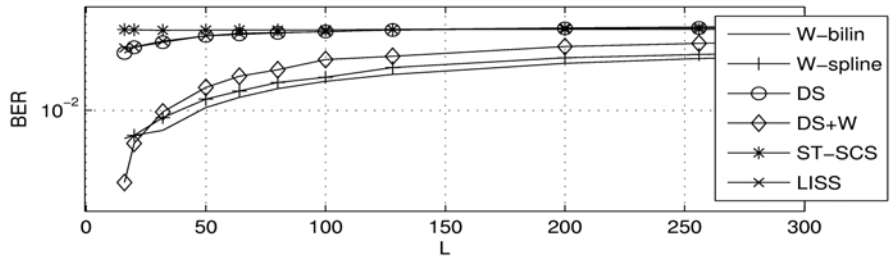

FIG. 4.37 – Robustesse de W-interp au débruitage (DNR=14 dB), DWR=28 dB

Attaques valumétriques

Dans ce paragraphe, on utilise le décodeur itératif présenté dans l'annexe C.2 pour W-interp. Si le gain est constant (cf. *fig.* 4.38), W-bilin et RDM sont invariantes à l'attaque (TEB=0), contrairement aux techniques quantificatives classiques. Lorsque le gain varie d'un pixel à l'autre selon une "marche aléatoire" $\rho_k = \rho_{k-1} + n_k$, où $\mathbf{n}$ est un AWGN, RDM avec un code court ($N_v = 4$) est pratiquement invariant, contrairement à RDM avec un code long ($N_v = 10$). W-bilin est sensible à l'attaque, mais moins que SCS et ST-SCS (cf. *fig.* 4.39). Dans les trois cas, on notera la meilleure robustesse de mod-LPTV-SCS sur ST-SCS.

W-interp (comme LISS et DS+W) est très robuste à l'égalisation d'histogramme, à laquelle ST-SCS est particulièrement sensible. De plus, le fait que W-interp rejette les interférences de l'hôte l'avantage par rapport à DS. RDM n'améliore que peu la robustesse de SCS à l'égalisation d'histogramme, ce qui est un défaut majeur de cette technique destinée à améliorer la robustesse aux attaques valumétriques. On constate à nouveau sur la *fig.* 4.40 que, dans le cas d'un DNR élevé, LISS offre de meilleures performances que W-interp pour L faible, mais qu'à partir de $L > 300$ (pour DWR=28 dB), W-interp devient meilleure.

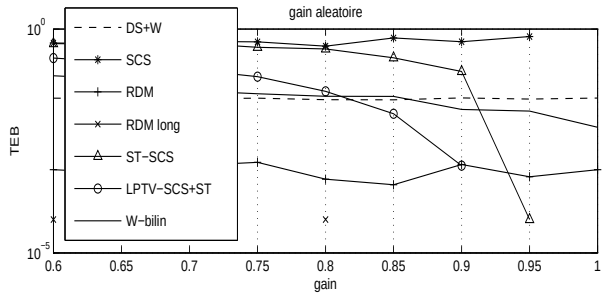

FIG. 4.38 – Robustesse à une attaque de gain constant, $L = 100$, DWR=28 dB

4.6.4 Attaques géométriques et bruit d'interpolation

Robustesse à une désynchronisation

W-interp, comme toutes les autres méthodes de cette comparaison, n'est pas robuste à une attaque géométrique telle que la rotation, même d'un angle faible (cf. *fig.* 4.41). W-interp est sensible aux deux source d'erreurs de l'attaque : d'abord, l'interpolation nuit à la détection. Ensuite, W-interp est sensible à la désynchronisation. Une solution classique pour lutter contre l'effet désynchronisant des attaques géométriques est

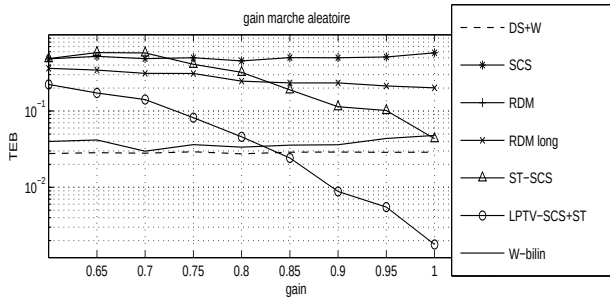

FIG. 4.39 – Robustesse à une attaque de gain variant selon une marche aléatoire, $L = 100$, DWR=28 dB

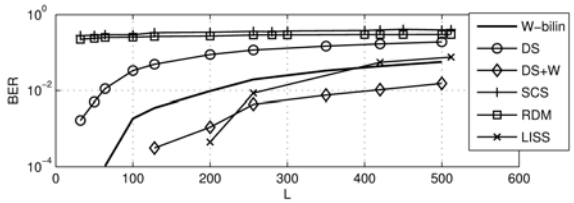

FIG. 4.40 – Robustesse à l'égalisation d'histogramme, W-interp, DWR=28 dB

fondée sur la resynchronisation (cf. paragraphe 1.5.4). Il est envisageable de combiner W-interp avec une méthode de resynchronisation. On peut par exemple superposer au tatouage de W-interp un tatouage additif à spectre étalé basé sur les pilotes. Un premier décodage est effectué sur les pilotes, puis le message est décodé par W-interp après resynchronisation. Dans la suite de cette section, on étudiera donc la robustesse de W-interp à un bruit d'interpolation, généré par exemple par un changement d'échelle ou une translation. On y suppose que la resynchronisation est parfaite grâce à l'utilisation de pilotes.

Les simulations montrent cependant que même après resynchronisation parfaite, le WNR du bruit d'interpolation est trop élevé pour W-interp. En effet, il n'est pas possible de recourir à un motif de référence comme pour DS, afin d'économiser une interpolation. La normalisation d'images présente les mêmes défauts que la resynchronisation (cf. paragraphe 1.5.4) : elle introduit un bruit d'interpolation. Notons cependant que ce bruit peut être connu lors de l'insertion. Ceci permet d'envisager une technique d'insertion informée (cf. paragraphe 4.3.2). W-interp est toujours particulièrement sensible aux attaques géométriques car elle n'est pas robuste à une attaque de très faible WNR (impossible avec les attaques classiques). L'annexe C.5 présente des pistes pour améliorer la robustesse de W-interp aux attaques géométriques.

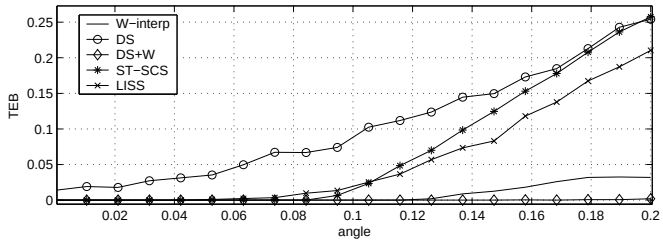

FIG. 4.41 – Robustesse à la rotation, W-interp, $L = 64$, DWR=28 dB

Robustesse au changement d'échelle

Cette attaque est particulièrement intéressante pour W-interp puisque l'erreur qu'elle introduit au décodage est due à une interpolation systématique. Plus δ est faible, plus la déformation est grande (et plus DNR est faible). $\delta = 0.5$ correspond à DNR=12 dB, ce qui est une attaque très perceptible. Une attaque imperceptible (DNR=26 dB) correspond à $\delta = 61/64$. W-interp est robuste à un changement d'échelle de faible facteur suivi d'une resynchronisation (cf. paragraphe *fig.*4.42). Lorsque δ est petit, W-interp est plus sensible à l'attaque que les autres techniques. Ceci est cohérent avec les performances face à l'attaque AWGN : ici, $\delta = 0.5$ correspond à WNR=-16 dB, pour lequel les performances théoriques de W-interp face au bruit AWGN sont mauvaises (cf. paragraphe 4.2.2).

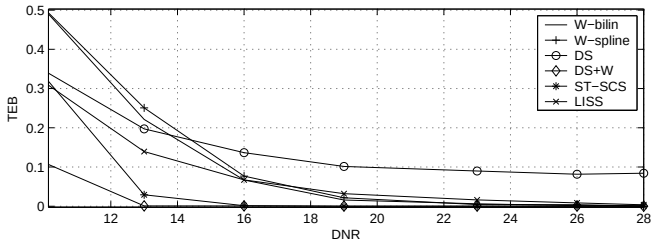

FIG. 4.42 – Robustesse au bruit d'interpolation, $L = 128$, DWR=28 dB

Robustesse à la translation et au rognage

Dans le tatouage par étalement de spectre classique, le maximum de la corrélation entre la séquence PN et le tatouage permet de retrouver les décalages (τ_h, τ_v). Ensuite il suffit d'effectuer la translation inverse. Si $(\tau_h, \tau_v) \in \mathbb{Z}^2$, la resynchronisation du tatouage est parfaite (on perd juste des échantillons si l'image a été rognée). Si les décalages ne sont pas entiers, un bruit d'interpolation apparaît, même après resynchronisation. Si l'on superpose à W-interp un tatouage pilote utilisant l'étalement de spectre, en partageant la puissance, la reconstruction est parfaite si $(\tau_h, \tau_v) \in \mathbb{Z}^2$. Sinon, la robustesse suit une fonction périodique de (τ_h, τ_v) (cf. *fig.* 4.43).

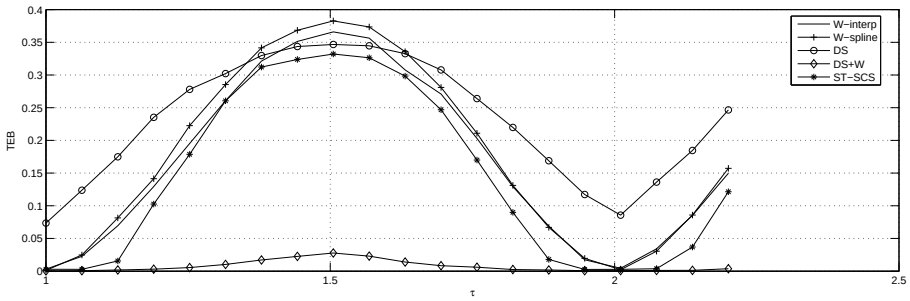

FIG. 4.43 – Robustesse à la translation, après resynchronisation (bruit d'interpolation)

4.6.5 Tableau récapitulatif de la robustesse

SCS et ST-SCS sont réputées fragiles aux attaques valumétriques (égalisation d'histogramme), mais les simulations indiquent également une faible robustesse aux autres attaques (débruitage, rotation…). Lorsque l'attaque est puissante, LISS perd

sa capacité de rejet des interférences de l'hôte et ses performances tombent à un niveau proche de DS (compression, débruitage). C'est la méthode DS+W (et donc LISS+W) qui est la plus robuste aux attaques. W-spline est moins robuste que W-interp à la compression JPEG. La vulnérabilité aux autres attaques est similaire. W-interp présente donc un compromis de robustesse intéressant. Elle est beaucoup moins robuste que les techniques quantificatives à un AWGN, mais plus robuste à des attaques telles que la compression, le débruitage et surtout l'égalisation d'histogramme (y compris par rapport à RDM). DS+W partage ces propriétés de robustesse, mais ne présente pas de rejet des interférences de l'hôte. C'est par rapport à LISS (et *a fortiori* LISS+W) que W-interp présente le moins d'intérêt.

	DS+W	LISS	ST-SCS	SCS	RDM	W-bilin	W-spline
AWGN faible	moyenne	selon L	excellente	excellente	excellente	très bonne	bonne
AWGN fort	très bonne	très bonne	excellente	très bonne	très bonne	moyenne	moyenne
débruitage	très bonne	moyenne	mauvaise	mauvaise	mauvaise	excellente	excellente
compression	excellente	mauvaise	moyenne	très bonne	très bonne	très bonne	bonne
gain constant	invariant		très mauvaise	très mauvaise	invariant	invariant	invariant
gain variable	très bonne		mauvaise	très mauvaise	excellente	très bonne	très bonne
ég. d'hist.	excellente	très bonne	très mauvaise	très mauvaise	mauvaise	très bonne	très bonne
bruit d'interp.	très bonne	moyenne	moyenne			mauvaise	mauvaise

4.7 Sécurité de W-interp

Dans ce paragraphe, le pirate veut estimer $K = \{\mathcal{S}, \mathcal{T}\}$. W-interp modifie la distribution de l'erreur d'interpolation $\epsilon(\mathbf{x})$. Cette distribution est calculée pour $\mathbf{g}$ donnée, mais une information sur la clé fuit même si l'on ne connaît pas exactement $\mathbf{g}$. Selon la terminologie de [CFF05], W-interp n'est donc pas à couverture parfaite. Un calcul du niveau de sécurité de W-interp est par conséquent nécessaire. Dans ce paragraphe, les niveaux de sécurité théoriques de W-interp sont étudiés dans le cadre établi par [CFF05]. Une attaque pratique sur la sécurité, spécifique à W-interp est ensuite proposée. Elle permet d'évaluer le niveau de sécurité empirique de W-interp lorsque plusieurs images tatouées avec la même clé sont disponibles. Les attaques à message connu (KMA) et à original connu (KOA) utilisent des versions simplifiées de cet algorithme.

4.7.1 Niveau de sécurité théorique

Il est possible de calculer le niveau de sécurité théorique de W-interp en exprimant la matrice d'information de Fisher ou en calculant l'équivoque de Shannon (cf. paragraphe 1.4.2). Dans ce dernier cas, la plupart des calculs d'intégrales ne peuvent se faire que numériquement. Dans l'approche de Fisher, les travaux réalisés pour DS ne sont pas transposables au cas de W-interp. La sécurité de W-interp ne repose pas sur une seule clé, mais sur le lien entre les deux parties distinctes de la clé : les fonctions d'interpolation $\mathbf{g}$ et la mise en forme $\mathcal{S}_1 \cup \cdots \cup \mathcal{S}_L$. Nos efforts pour exprimer le niveau de sécurité théorique de W-interp n'ont pas abouti : nous sommes amenés à effectuer des simplifications abusives qui ramènent la sécurité de W-interp à un comportement proche de celui de DS. Comme nous le montrons dans la suite, le niveau de sécurité de W-interp est pourtant clairement inférieur à celui des techniques DS. Dans la suite,

nous proposons donc une approche plus *ad-hoc* de la sécurité de W-interp.

Le modèle d'insertion est :

$$y_k = x_k + b_k(g_k(\underline{x}_k) - x_k) = x_k + b_k(\sum_{j=1}^{N_v} g_j x_k^j - x_k)$$

Donc si le pirate possède N_v échantillons où $b_k = 1$, on a $\mathbf{g}[\underline{x}_k^1, \ldots, \underline{x}_k^{N_v}]^{\mathrm{T}} = [y_k^1, \ldots, y_k^{N_v}]$. Le calcul de $\mathbf{g}$ revient donc à la simple inversion d'un système linéaire. W-spline offrira un meilleur niveau de sécurité car N_v est plus grand (en théorie, le support est même infini).

La sécurité à une attaque avec $N_o > 1$ ne peut donc reposer que sur le bruitage de g. Si on ajoute un bruit, il y a un compromis entre l'augmentation de la sécurité et la diminution des performances au décodage. Le modèle d'insertion devient alors pour la sécurité :

$$y_k = x_k + b_k(g_k(\underline{x}_k) + n_k - x_k) \tag{4.4}$$

ou encore

$$y_k = x_k + Q(b_k(g_k(\underline{x}_k) - x_k))$$

Le niveau de sécurité dépend de $\sigma^2_{\mathbf{n}}$. Dans le cas de base, $\sigma^2_{\mathbf{n}} = 1/12$ à cause de la quantification de codage de l'image. Pour augmenter la sécurité, on pourrait également rajouter un bruit plus fort dans l'hypothèse H_1, soit sur l'échantillon k, soit sur tout le document. Cela aurait peu d'impact sur le décodage (changement de la variance dans l'hypothèse H_1 ou ajout de $\epsilon(\mathbf{n})$) tant que WNR est proche de 1. Il y aurait par contre un impact perceptuel. Si l'on insère une quantification, on ne rajoute pas de bruit au décodage mais on perd la linéarité (donc l'invariance aux attaques valumétriques).

Une autre solution pour augmenter la sécurité de W-interp consiste à utiliser une interpolation $\mathbf{g}$ non linéaire. Au bilan, le niveau de sécurité de W-interp est très élevé si $N_o = 1$ (il résulte d'un compromis avec l'imperceptibilité) mais faible si $N_o > 1$. Il résulte alors d'un compromis avec les performances au décodage.

Enfin, la compensation des distorsions (cf. paragraphe 4.3.2) permet d'améliorer le niveau de sécurité. En effet, comme x_k est inconnu du pirate, DC-W-interp revient à introduire le bruit $n_k = m_l(1-\alpha)(\tilde{x}_k - x_k)$. Cependant, ceci se fait au détriment des performances au décodage si l'introduction de $\alpha < 1$ n'est pas justifiée par la présence d'une attaque AWGN. Ce comportement est similaire à celui observé pour les techniques quantificatives (cf. paragraphe 1.4.2).

4.7.2 Algorithmes pratiques d'attaques sur la sécurité spécifiques à W-interp : KMA et KOA

Un algorithme pratique d'attaque sur la sécurité a trois objectifs :

1) estimer $\mathcal{S}$
2) estimer $\{\mathcal{S}_l, l \in \{1, \ldots, L\}\}$
3) estimer $\mathbf{g}$

Si $N_o = 1$, les trois estimations doivent être conjointes. Si $N_o > 1$, nous proposons une stratégie sous-optimale pour simplifier la construction de l'algorithme d'attaque. Elle consiste à estimer g^k indépendamment pour chaque échantillon, et à combiner ces estimations pour estimer $\mathcal{S}$. On notera $\mathcal{S}_{+1}^{n_o} = \cup_{l=1,\ldots,L|m_l^{n_o}=+1}\mathcal{S}_l$.

Cas où $N_o = 1$

Attaque KOA : le pirate a accès à $\mathbf{w} = \mathbf{b}(\mathbf{g}(\underline{\mathbf{x}})+\mathbf{n}-\mathbf{x})$, à $\underline{\mathbf{y}} = \underline{\mathbf{x}}$ et à $\mathbf{x}$. S'il décide que $w_k = 0$ implique que $b_k = 0$, il se ramène à $\mathbf{w} = \mathbf{g}(\underline{\mathbf{x}}) + \mathbf{n}$. Il n'a d'autre solution que de supposer (à tort) que $\mathbf{g}$ est constant, et se ramène à la résolution du système linéaire bruité : $\mathbf{w} = [g_j]\underline{\mathbf{x}}+\mathbf{n}$. La solution d'un système linéaire bruité $Y = \Theta X + N$, en connaissant Y et X, est estimée par $\hat{\Theta} = YC$, où $C = X^\top(XX^\top)^{-1}$ sans biais et de variance $C^\top C\sigma_N^2$. Comme $\sigma_{\mathbf{n}}^2$ est faible, g serait estimé correctement si $\mathbf{g}$ était constant.

Attaques WOA et KMA : désormais le pirate n'a accès qu'à $\mathbf{y} = \mathbf{x} + \mathbf{b}(\mathbf{g}(\underline{\mathbf{x}}) + \mathbf{n} - \mathbf{x})$ et $\underline{\mathbf{y}} = \underline{\mathbf{x}}$. La résolution du système linéaire bruité sur $k \in \{1, \ldots, N\}$ sous l'hypothèse $\mathbf{g} = g$ constant est très sous-optimale. Le pirate doit donc estimer $\mathcal{S}_{+1}$ simultanément à g.

Dans les trois cas, l'estimation de $\{\mathcal{S}_l, l \in \{1, \ldots, L\}\}$ est impossible. Comme la mise en forme est secrète, la connaissance de $\mathbf{m}$ dans KMA est inutile ($\mathbf{b}$ reste inconnu).

Cas où $N_o > 1$

Attaque KOA : à la coordonnée k, le pirate a accès à $[w_k^{n_o}] = [b_k^{n_o}](g^k([\underline{x}_k^{n_o}]) + [n_k^{n_o}] - [x_k^{n_o}])$, à $[\underline{y}_k^{n_o}] = [\underline{x}_k^{n_o}]$ et à $[x_k^{n_o}]$. On se ramène au cas $N_o = 1$, mais cette fois-ci l'hypothèse $\mathbf{g} = g$ constant est correcte. La variance de l'estimateur est $CC^\top\sigma_{\mathbf{n}}^2$, où $C = [\underline{x}_k^{n_o}]^\top([\underline{x}_k^{n_o}][\underline{x}_k^{n_o}]^\top)^{-1}$. W-interp n'est pas sûr à cette attaque.

Attaques WOA et KMA : comme précédemment, le pirate doit estimer $\mathcal{S}_{+1}^{n_o}$ simultanément à $\mathbf{g}$. En cas d'estimation parfaite de $\mathcal{S}_{+1}^{n_o}$, la variance de l'estimateur de $\mathbf{g}$ est $CC^\top\sigma_{\mathbf{n}}^2$. Le niveau de sécurité de W-interp dépend donc de la difficulté pratique de l'estimation simultanée.

La connaissance (avec KOA) ou l'estimation (WOA et KMA) de $\{\mathcal{S}_{+1}^{n_o}, n_o = 1, \ldots, N_o\}$ permet ensuite d'estimer $\{\mathcal{S}_l, l \in \{1, \ldots, L\}\}$ par un algorithme de séparation de source [CFF05], puis de décoder $\mathcal{M}$. Un algorithme plus complexe consisterait à estimer $\{\mathcal{S}_l, l \in \{1, \ldots, L\}\}$ simultanément à $\mathbf{g}$ et $\{\mathcal{S}_{+1}^{n_o}, n_o = 1, \ldots, N_o\}$. On serait alors en mesure d'exploiter la connaissance de $\mathcal{M}$ dans KMA.

4.7.3 Un algorithme EM pour WOA

Principe de l'algorithme proposé pour WOA

Dans un cadre très différent de W-interp, Popescu et Farid [PF05] ont développé une technique de détection des altérations d'une image numérique générée par un appareil utilisant des *Color Filter Array* (CFA), tel qu'un appareil photo numérique. Les images étudiées sont générées par des filtres utilisant diverses matrices d'interpolation. Les auteurs utilisent un algorithme d'Espérance-Maximisation (EM) [DLR77] pour maximiser la vraisemblance des coefficients spécifiques du filtre d'interpolation tout en estimant les pixels interpolés. On peut alors exhiber des structures de corrélation spécifiques dans le domaine de Fourier de la carte de probabilité produite par l'algorithme EM. Si l'image est manipulée à l'aide d'un logiciel tel que Photoshop, ces corrélations spécifiques disparaissent à l'emplacement modifié. On peut donc détecter à la fois la modification et son emplacement. Il s'agit donc d'une technique alternative au tatouage fragile, sans insertion de tatouage.

Cet algorithme peut servir à construire une attaque spécifique contre W-interp. En effet, la sécurité de W-interp repose à la fois sur le secret des points interpolés et sur celui des paramètres de l'interpolation employés. L'algorithme EM proposé par Popescu *et al.* permet d'estimer les deux simultanément. [PF05] montre qu'il est efficace lorsque les paramètres d'interpolation sont constants sur toute l'image. Il convient de vérifier expérimentalement si son efficacité subsiste lorsque les paramètres varient pour chaque pixel interpolé, et lorsque le pirate a accès à plusieurs images tatouées avec la même clé. L'algorithme pourrait également servir à différencier deux variantes de W-interp, bien que selon les principes de Kerckhoff, la sécurité ne doive pas s'appuyer sur le secret de la variante employée.

Présentation de l'algorithme EM utilisé

L'algorithme EM [DLR77] est une méthode d'optimisation itérative qui consiste à estimer des paramètres Θ, à partir de données $\mathbf{x}$. De plus, une inconnue supplémentaire gêne l'estimation : les paramètres $\mathcal{P}$. On veut donc maximiser Θ en fonction de la distribution de $\mathcal{P}$ et à partir de $\mathbf{x}$:

$$\Theta^* = \text{argmax}_\Theta \sum_{\mathcal{P}} \text{p}(\Theta, \mathcal{P}|\mathbf{x}) \tag{4.5}$$

Le principe de l'algorithme EM est alors d'alterner une étape d'estimation de $\mathcal{P}$ pour Θ données, puis une étape de maximisation où l'on choisit Θ en fonction de l'estimation de $\mathcal{P}$ précédente. En sortie de l'algorithme, on a donc une estimation Θ^* ainsi que la vraisemblance des paramètres inconnus $\mathcal{P}$. L'étape d'estimation calcule une borne inférieure locale de la distribution *a posteriori*, alors que l'étape de maximisation optimise cette borne et améliore l'estimation de Θ. Une preuve de convergence de l'algorithme est possible [DLR77]. L'étape d'estimation à l'itération t se fait selon une distribution de $\mathcal{P}$ sachant Θ^t : $\text{p}(\mathcal{P}|\mathbf{x}, \Theta^t)$. On a donc une décision souple (probabilité *a posteriori*) de $\mathcal{P}$. Dans l'étape de maximisation, on optimise la borne $\log(\text{p}(\mathbf{x}, \mathcal{P}|\Theta)) + \text{p}(\Theta)$, somme de la log-vraisemblance et de la probabilité a priori de Θ.

Ici, $\Theta = \{g_j, j \in \{1, \ldots, N_v\}\}$ et $\mathcal{P} = \mathcal{S}$. La distribution *a priori* de $\epsilon(y_k)$ est calculée d'après la règle de Bayes à partir des probabilités *a priori* d'appartenir à $\mathcal{S}$ ou non et d'une distribution de l'erreur d'interpolation sur chaque classe. L'optimisation de Θ minimise l'erreur quadratique par la méthode des moindres carrés, pondérés par la probabilité *a posteriori* de $\mathcal{P}$. Cette minimisation consiste à annuler les dérivées partielles de l'erreur quadratique selon chaque g_j, donc à résoudre un système linéaire [PF05].

Implantation de l'algorithme WOA et adaptations

Dans l'implantation, les adaptations suivantes ont été apportées à l'algorithme de Popescu et Farid :

- La deuxième partie de l'algorithme de [PF05] consiste à effectuer une FFT pour détecter une périodicité (due aux capteurs de l'appareil) dans la carte de probabilité. Elle est inutile ici.
- dans le cas H_0 (non tatoué), on modélise l'erreur d'interpolation par une GGD($c_{\epsilon(\mathbf{x})}$,$\sigma^2_{\epsilon(\mathbf{x})}$) au lieu d'une distribution uniforme.
- au lieu d'une équiprobabilité des cas H_0 et H_1, on utilise $p_{H_1} = \frac{1}{2}\frac{2}{N}N_\mathcal{S} = \frac{\sigma^2_\mathbf{x}}{\sigma^2_{\epsilon(\mathbf{x})}\text{DWR}}$.

ALGORITHME EM
/* Initialisation */
$t = 0$ // itération
$\sigma_t = 1$
$p_{H_1} = \frac{\sigma_{\mathbf{x}}^2}{\sigma_{\epsilon(\mathbf{x})}^2 \text{DWR}}$
$N_v = 1$ // pour W-bilin
tant que $\sum_{j=1}^{N_v} |(g_j)^{(t)} - (g_j)^{(t-1)}| < \epsilon$
/* étape d'estimation */
pour toutes les coordonnées k
$\epsilon(x_k) = |y_k - \sum_{j=1}^{N_v} \alpha_j^{(t)} \underline{y}_{k,j}|$
fin
pour toutes les coordonnées k
$f_{H_1}(k) = \frac{1}{\sigma_t\sqrt{2\pi}} e^{-\epsilon^2(x_k)/2\sigma_t^2}$
$f_{H_0}(k) = Ae^{-|\beta\epsilon(x_k)|^c}$
$w_k = \frac{p_{H_1} f_{H_1}(k)}{(1-p_{H_1}) f_{H_0}(k) + p_{H_1} f_{H_1}(k)}$
fin
/* étape de maximisation */
calculer g solution d'un système linéaire [PF05]
$\sigma_{t+1} = \sqrt{\frac{\sum_k w_k \epsilon^2(x_k)}{\sum_k w_k}}$
$t = t + 1$
fin

- à l'initialisation, $g_j = \frac{1}{N_v} \quad \forall j$.
- pour initialiser σ, on calcule (cf. annexe C.4) l'erreur moyenne entre l'interpolation de base et une réalisation de g : $\sigma_{\text{EM}}^2 = \frac{63}{144}\sigma_{\epsilon(\mathbf{x})}^2$.
- on choisit comme critères d'arrêt $\epsilon = 0.001$ et $N_{\text{it}} = 100$.

Deux attaques sont envisagées. La première attaque consiste à appliquer l'algorithme EM sur **un seul document tatoué** (N_o=1), afin de déterminer $\{\cup\mathcal{S}_l | m_l = +1\}$, ainsi que $\mathbf{g}$ sur cet ensemble. Trois scénarios sont envisagés : $\mathbf{g}$ de base, $\mathbf{g}$ constant, $\mathbf{g}$ variable sur chaque pixel (correspondant à W-interp). Lorsque DWR diminue, le cardinal de $\{\cup\mathcal{S}_l | m_l = +1\}$ augmente et l'estimation est plus facile. A DWR donné, on estime $\{\cup\mathcal{S}_l | m_l = +1\}$ comme les échantillons correspondant aux $N_{\mathcal{S}}/2$ plus grandes valeurs de la carte de probabilité.

L'étude précédente montre qu'un seul document ne suffit pas à estimer $\mathbf{k}$ pour W-interp. Lorsque le pirate dispose de $N_o > 1$ **documents distincts tatoués avec la même clé**, il peut se ramener à l'attaque précédente dans le scénario $\mathbf{g}$ constant. Pour chaque point k, l'algorithme EM est alors appliqué au signal

$$\{\underline{y}_k^{n_o}, \quad k_o \in \{1, \ldots, N_o\}\}$$

Si $k \in \mathcal{S}$ et si l'algorithme converge (donc si N_o est assez grand), on estime $\{g_j^k, j \in \{1, \ldots, N_v\}\}$. Si $k \notin \mathcal{S}$, l'algorithme n'est pas supposé converger, sauf propriétés particulières des images. Pour chaque point k, on décide donc $k \in \mathcal{S}$ si $\sigma_{\text{EM}}^2 < \nu_{\sigma_{\text{EM}}^2}$ où $\nu_{\sigma_{\text{EM}}^2}$ est un seuil donné. De plus, si $k \in \mathcal{S}$, comme $\mathbf{m}^{n_o}$ varie pour chaque image, seule la moitié des éléments de $\{y_k^{n_o}, \quad n_o \in \{1, \ldots, N_o\}\}$ est le résultat d'une substitution.

La carte de probabilité permet donc d'estimer $b_k^{n_o}$, bit inséré en k. Puis m_l peut être estimé si $\mathcal{S}_l$ est estimé à partir des $\{b_k^{n_o}, \quad n_o \in \{1,\dots,N_o\}, \quad k \in \{1,\dots,N\}\}$.

Application à l'image : conditions expérimentales

Les simulations ont été effectuées sur W-bilin pour des raisons de simplicité. Elles peuvent être étendues à W-spline (avec une approximation du support à 4x4), ainsi qu'à d'autres variantes de W-interp. La base d'images utilisées pour le test est celle de la *City University of Hong Kong* [Cit]. Sur cette base, $0.27 < c < 1.7$, avec en moyenne $c = 0.6$. De même, $4.6 < \sigma_{\epsilon(\mathbf{x})} < 17.5$, avec en moyenne $\sigma_{\epsilon(\mathbf{x})} = 10.4$. Une image non tatouée possède déjà des propriétés d'intercorrélation entre pixels voisins, qui gênent l'estimation et la classification par l'algorithme EM. Dans [PF05], l'algorithme est performant car la probabilité pour qu'un point soit interpolé est de $1/2$. Dans le cas du tatouage, p_{H_1} peut être beaucoup plus faible (en fonction du DWR), ce qui augmente la sécurité et gêne l'attaque.

Résultats expérimentaux de l'attaque WOA, $N_o = 1$

La *fig.* 4.44 illustre l'intérêt des cartes de probabilités produites par l'algorithme EM. Elle présente les probabilités pour chaque point de l'image d'être tatoué, dans un cas où l'algorithme converge (ici, $\mathbf{g}$ constant). Les points les plus sombres correspondent aux probabilités les plus élevées. Lorsque l'image est non tatouée, on retrouve les caractéristiques de l'image : les points où $\epsilon(y_k)$ est la plus grande, donc où la probabilité d'être tatoué est la plus faible, sont situés sur les contours. Lorsque l'image est fortement tatouée, les points les plus probablement tatoués sont à l'inverse disséminés sur toute l'image (contours compris).

La *fig.* 4.45 présente le résultat de l'attaque sur la sécurité lorsque $N_o = 1$. Si $\mathbf{g}$ est constant, l'algorithme EM réussit à estimer parfaitement $\mathbf{g}$ dès que DWR est suffisamment faible (DWR<24 dB). De l'estimation de $\mathbf{g}$, on peut déduire les décalages utilisés pour la sécurité de W-bilin. Si DWR est grand, les propriétés propres à l'image de corrélation locale entre pixels voisins sont prépondérantes et l'attaque échoue toujours. De même, on peut estimer correctement la position de jusqu'à 80% des points de $\mathcal{S}$. Lorsque $\mathbf{g}$ varie pour chaque pixel tatoué, l'estimation échoue pour toutes les valeurs du DWR. On peut estimer correctement la position de jusqu'à 25% des points de $\mathcal{S}$ à DWR=22 dB. Cependant, la proportion de points tatoués ($N_{\mathcal{S}}/2N$) augmente lorsque DWR diminue et ce pourcentage est à peine supérieur à un choix aléatoire des points. $\mathbf{g}$ garantit donc la sécurité de W-interp lorsque $N_o = 1$.

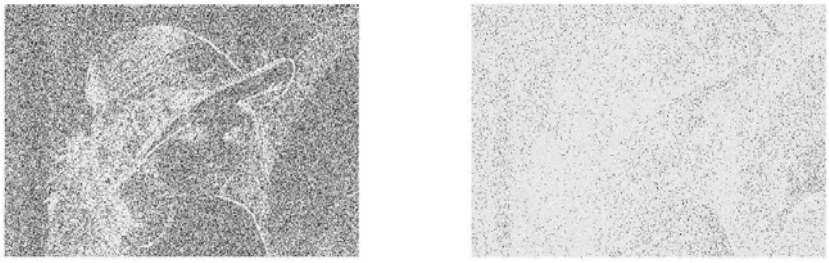

FIG. 4.44 – Cartes de probabilités, Lena, $\mathbf{g}$ constant, (gauche) non tatouée, (droite) tatouée, DWR=22 dB

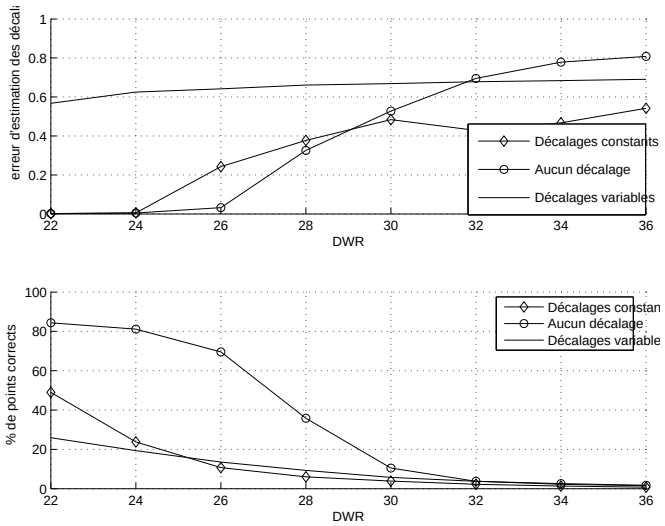

FIG. 4.45 – Attaque sur la sécurité, $N_o = 1$, W-bilin

Résultats expérimentaux de l'attaque WOA, $N_o > 1$

La *fig.* 4.46 présente les résultats de l'attaque lorsque $N_o > 1$ et $\mathbf{g}$ varie. $\mathbf{g}$ n'est estimé précisément que si N_o est très grand. Cependant, σ^2_{EM} converge rapidement même s'il y a une légère erreur d'estimation sur $\mathbf{g}$. Pour $N_o > 1000$ images, on peut estimer correctement la position des points de $\mathcal{S}$. Si $N_o < 1000$, l'étude met en évidence la sécurité de W-bilin. Même lorsque N_o est grand, TEB$\simeq 0.2$, donc l'attaque ne réussit pas parfaitement. La *fig.* 4.47 montre σ^2_{EM}, illustrant la convergence de l'algorithme. Lorsque N_o augmente, σ^2_{EM} converge sur $\mathcal{S}$ vers une valeur limite (en théorie, correspondant à $\sigma^2_{\mathbf{n}}$; en pratique l'erreur est plus grande) et augmente sur $\{1, N\} \setminus (\mathcal{G} \cup \mathcal{S})$. Cette étude confirme donc la sécurité de W-interp, dans la plupart des applications du tatouage numérique. Notons cependant que dans certains cas, comme le tatouage vidéo, un grand nombre d'images est disponible, donc $N_o > 1000$ peut être atteint. De plus, le niveau de sécurité pratique de DS est meilleur [CFF05].

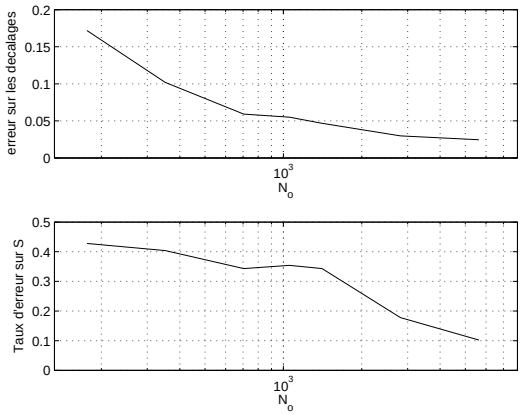

FIG. 4.46 – Attaque sur la sécurité, $N_o > 1$, W-bilin

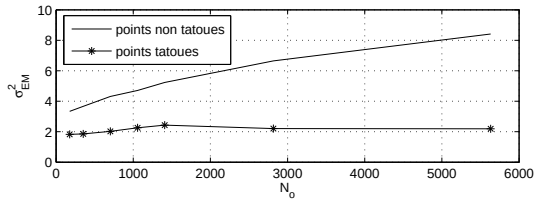

FIG. 4.47 – Attaque sur la sécurité, $N_o > 1$, W-bilin : convergence de l'algorithme EM

Attaque "intelligente" sur la robustesse associée à l'attaque WOA

Après attaque sur la sécurité, on peut utiliser l'information sur **k** pour construire une attaque sur la robustesse spécifique [CFF04]. Ici, on ajoute un bruit AWGN sur l'estimation de $\mathcal{S}$. La *fig.* 4.48 montre que la combinaison des deux attaques est très efficace (avec le bruit additif gaussien seul, le TEB est toujours nul). D'autres attaques sont étudiées dans l'annexe C.4.

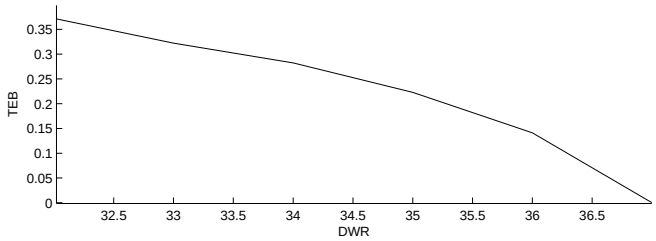

FIG. 4.48 – Attaque sur la robustesse combinée avec l'estimation de la clé, $L = 256$, WNR=1, cas $N_o = 1$

4.7.4 Conclusion et tableau récapitulatif

On a prouvé théoriquement et expérimentalement la sécurité de W-interp lorsque $N_o = 1$. Lorsque $N_o > 1$, le niveau de sécurité de W-interp est faible. Cependant, l'application d'algorithmes d'attaque de type WOA pratiques ne réussit que pour $N_o > 1000$, niveau de sécurité acceptable.

Plusieurs pistes sont possibles pour améliorer la sécurité de W-interp. On pourrait introduire une part d'aléatoire (donc de secret) dans le choix de $\mathcal{G}$, pour compliquer l'accès à $\underline{y}$. Notamment, il serait intéressant d'étudier l'utilisation de grilles non équiréparties. L'utilisation de plusieurs grilles d'interpolation imbriquées est également possible. La déformation sera cependant plus importante (**g** s'éloigne de la fonction d'interpolation de base), ainsi que le coût calculatoire. On pourrait également utiliser deux fonctions d'interpolation $g^k|b_k = 0$ et $g^k|b_k = +1$. Cette technique se rapproche cependant d'une technique RDM classique.

	KOA	KMA	WOA
DS	BSS sans bruit	résolution syst. linéaire bruité	BSS avec bruit
PCC	idem DS + entropie de **k** différente + reconnaissance de code		
LPTV	idem DS + entropie de **k** différente + changement de domaine		
W-interp	résolution syst. linéaire bruité, sécurité faible	algorithme EM, sécurité pratique élevée	algorithme EM, sécurité pratique élevée

4.8 Conclusion et extensions possibles

Dans ce chapitre, nous avons proposé une famille d'algorithmes de tatouage nommée W-interp. Nous avons étudié ses performances, sa robustesse, sa sécurité et son imperceptibilité. W-interp est un algorithme de tatouage aveugle et informé. Il peut être rapproché des techniques de catégorisation aléatoire, mais présente de nombreux points originaux. W-interp utilise les propriétés perceptuelles de l'interpolation, ce qui n'avait jamais été le cas jusqu'ici dans le domaine du tatouage, et qui permet de se passer de masque psychovisuel. W-interp a été proposé et analysé dans un cadre générique. Divers types de documents et diverses techniques d'interpolation sont envisageables. Parfois, la linéarité de la technique d'interpolation est exigée. Dans l'étude expérimentales, nous avons étudié deux cas particuliers simples en tatouage d'images : W-bilin, utilisant l'interpolation bilinéaire, et W-spline, utilisant les B-splines cubiques.

Nous avons souligné les liens de W-interp avec les techniques de quantification proportionnelle à l'hôte. En effet, W-interp peut être considérée comme une technique de catégorisation aléatoire. Elle comporte de nombreux éléments originaux, dont l'utilisation de "quantificateurs" linéaires, continus et non disjoints (à l'encontre des principes de constructions de codes de tatouage classiques). Le dictionnaire ainsi construit doit être transmis dans le document tatoué, ce qui nuit au débit accessible. D'autre part, le décodage par distance minimale ne peut être utilisé. En effet, nous n'avons pas cherché à rendre maximale la distance entre les sous-dictionnaires de W-interp, contrairement à l'approche classique. La combinaison de l'algorithme RDM avec un masque perceptuel fondé sur l'interpolation fournit donc une meilleure robustesse au bruit AWGN. Nous avons proposé une règle de décodage originale, qui utilise la redondance de façon intrinsèque. W-interp est donc fondamentalement une technique de tatouage à bas débit. Cependant, nous avons montré en pratique que la technique proposée a l'intérêt de fournir un masque perceptuel intrinsèque et une robustesse à de nombreuses attaques, y compris valumétriques. Ceci est un net avantage par rapport aux techniques de catégorisation aléatoire classiques.

L'imperceptibilité de W-interp provient de l'utilisation de l'interpolation. Elle est contrôlée par un critère de distorsion globale. Les mesures objectives de qualité perceptuelle ont confirmé les bonnes propriétés de W-bilin et W-spline. Cependant, nous avons montré qu'une étude subjective reste nécessaire. Celle-ci montre une meilleure imperceptibilité de W-spline, mais reste difficile à conduire de façon rigoureuse. Une étude perceptuelle subjective précise permettrait d'identifier les techniques d'interpolation les plus adaptées à l'application au tatouage.

La détection utilise un seuil empirique et adapté au contenu. Dans le scénario où l'attaque est connue à l'insertion, l'étude théorique permet de calculer un seuil de détection théorique qui améliore les performances de façon significative. Cela a été effectué pour l'attaque AWGN, et pourrait être étendu à d'autres attaques. D'autre part, nous avons proposé des extensions à l'insertion informée, et notamment à la compensation des distorsions. Les simulations ont montré une très bonne robustesse de W-interp aux attaques courantes agissant sur le signal : bruit additif indépendant de l'image, compression JPEG ou encore égalisation d'histogramme. W-interp présente également une meilleure robustesse au débruitage que les algorithmes de tatouage classiques étudiés car le tatouage inséré est très corrélé avec l'image. Cependant, W-interp est peu robuste aux attaques désynchronisantes, comme la plupart des algorithmes de tatouage. Des méthodes de resynchronisation spécifiques doivent être envisagées. Le problème est néanmoins plus complexe que pour les algorithmes de type DS car W-interp n'est

pas robuste à un ajout de bruit fort (au-delà du seuil d'imperceptibilité), qu'introduisent les méthodes de resynchronisation classiques.

Le niveau de sécurité de W-interp est garanti par le secret des points tatoués et l'introduction de décalages aléatoires, dont le rôle est similaire au signal d'agitation utilisé dans les techniques quantificatives. Des algorithmes pratiques d'attaque sur la sécurité de W-interp ont été proposés. Ils montrent que W-interp est sûre pour une seule image. Si plusieurs images tatouées avec la même clé sont en possession du pirate, le niveau de sécurité de W-interp est nettement inférieur à celui des techniques DS, mais une estimation précise de la clé reste difficile en pratique. L'utilisation de la compensation des distorsions permet d'améliorer le niveau de sécurité.

L'aspect générique de W-interp offre de nombreuses possibilités d'amélioration. Notamment, d'autres techniques d'interpolation évoquées dans la partie 3.1 sont prometteuses. Ainsi, l'utilisation de noyaux à symétrie radiale pourrait être envisagée pour augmenter la robustesse à des rotations locales. Les techniques d'interpolation préservant les contours sont de bons candidats pour améliorer la qualité perceptuelle subjective. Enfin, l'utilisation d'une technique d'interpolation non linéaire modifierait considérablement les propriétés de W-interp, et pourrait notamment améliorer le niveau de sécurité. On n'a envisagé en pratique que des grilles d'interpolation équiréparties. Le bénéfice d'autres voisinages $\underline{\mathbf{x}}$, par exemple choisis sur les contours, est à étudier. W-interp pourrait être adapté aux splines d'approximation, où la contrainte d'interpolation est relâchée au profit du débruitage, puisque le problème de la sécurité nous a conduit à modifier la fonction interpolante.

Nous nous sommes limités à une insertion dans le domaine spatial pour des raisons perceptuelles. Les algorithmes de tatouage classiques sont souvent plus robustes aux attaques dans le domaine transformé. L'application de W-interp à un domaine transformé est à donc étudier. Enfin, W-interp a été présenté et analysé théoriquement indépendamment du type de document. L'étude pratique s'est concentrée sur l'image. Il serait cependant intéressant d'étudier ses propriétés d'imperceptibilité et de robustesse pour d'autres types de documents où l'interpolation est utilisée : sons, vidéos.

Bibliographie

[Bas05] P. Bas. A quantization watermarking technique robust to linear and non-linear valumetric distortions using a fractal set of quantizers. *Information Hiding Workshop, Proc.*, pages 83–93, 2005.

[CFF04] F. Cayre, C. Fontaine, and T. Furon. Security of wss techniques. *Proc. Int. Workshop on Digital Watermarking (IWDW)*, 2004.

[CFF05] F. Cayre, C. Fontaine, and T. Furon. Watermarking security : Theory and practice. *IEEE Trans. on Signal Processing, Special Issue on Content Protection*, 53(10) :3976–3975, 2005.

[Cit] City University of Hong Kong Corel Image Database. http ://abacus.ee.cityu.edu.hk/ benjiman/corel_1/.

[CP95] J.-P. Cocquerez and S. Philipp. *Analyse d'images : filtrage et segmentation*. Masson, 1995.

[CW01] B. Chen and G.W. Wornell. Quantization index modulation : A class of provably good methods for digital watermarking and information embedding. *IEEE Trans. on Information Theory*, pages 1423–1443, 2001.

[DFHS03] J. Delhumeau, T. Furon, N. Hurley, and G. Silvestre. Improved polynomial detectors for side-informed watermarking. *Proc. SPIE*, 2003.

[DLR77] A. Dempster, N. Laird, and D. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, 99(1) :1–38, 1977.

[EG01] J.J. Eggers and B. Girod. Quantization effects on digital watermarks. *EURASIP Signal Processing*, 81(2) :239–263, 2001.

[FKK04] C. Fei, D. Kundur, and R.H. Kwong. Analysis and Design of Watermarking Algorithms for Improved Resistance to Compression. *IEEE Trans. on Image Processing*, 13(2) :126–144, 2004.

[Gre02] M.L. Green. Statistics of images, the TV algorithm of Rudin-Osher-Fatemi for image denoising and an improved denoising algorithm. *CAM reports, Univ. California, Los Angeles [Online] : http ://www.math.ucla.edu/applied/cam/index.html*, 2002.

[MCB00] M.L. Miller, I.J. Cox, and J.A. Bloom. Informed embedding : Exploiting image and detector information during watermark insertion. *IEEE Int. Conf. on Image Processing - ICIP*, 3 :1–4, 2000.

[MK05] P. Moulin and R. Koetter. Data-hiding codes. *Proc. of the IEEE*, 93(12) :2083–2127, 2005.

[PF05] A.C. Popescu and H. Farid. Exposing digital forgeries in color filter array interpolated images. *IEEE Trans. on Signal Processing*, 53(10) :3948 – 3959, 2005.

[PFCTPPG06] L. Pérez-Freire, P. Comesaña, J.R. Troncoso-Pastoriza, and F. Pérez-González. Watermarking security : a survey. *LNCS Transactions on Data Hiding and Multimedia Security. To appear.*, 2006.

[Sch00] H. Scharr. Optimal separable interpolation of color images with bayer array format. *Technical report, DFG research unit Image Sequence Analysis to Investigate Dynamic Processes*, 2000.

[SMCM05] K. Sullivan, U. Madhow, S. Chandrasekaran, and B. S. Manjunath. Steganalysis of spread spectrum data hiding exploiting cover memory. *Proc. SPIE*, pages 38–46, 2005.

[TBU00] P. Thévenaz, T. Blu, and M. Unser. Image interpolation and resampling. In I. Bankman, editor, *Handbook of Medical Imaging, Processing and Analysis*, chapter 25, pages 393–420. Acad. Press, San Diego, USA, 2000.

Chapitre 5

Conclusion

La première contribution de cette thèse a été d'appliquer les changements d'horloge périodiques (PCC) et les filtres linéaires variant périodiquement dans le temps (LPTV) au tatouage numérique par étalement de spectre. La substitution d'un étalement par filtre LPTV à l'étalement classique par code a été étudiée dans l'ensemble des techniques de tatouage impliquant un étalement. L'étude théorique a montré qu'avec les hypothèses classiques (signaux et attaques gaussiens), l'étalement par filtre LPTV est une alternative qui possède les mêmes performances que l'étalement DS classique. De plus, les filtres LPTV envisagés sont très simples de principe et d'implantation. Trois techniques d'étalement ont été privilégiées : les PCC utilisant des permutations aléatoires, les filtres LPTV sans perte et les filtres LPTV inversibles à filtres modulateurs constants. La sécurité des filtres LPTV a été étudiée et un algorithme pratique d'attaque sur la sécurité a été proposé. Les filtres LPTV obéissent aux mêmes principes de sécurité que la technique DS, la différence résidant dans la structure de la clé utilisée et le domaine d'insertion.

Les techniques proposées ont été appliquées en pratique au tatouage d'images numériques "naturelles". Cette étude expérimentale a mis en avant l'intérêt de l'utilisation d'un parcours d'image, qui permet d'exploiter les propriétés de corrélation spatiale d'une image naturelle. Cette propriété est habituellement contournée : les techniques de type DS utilisent souvent une mise en forme aléatoire qui décorrèle l'image. Nous avons montré que les PCC et les filtres LPTV sont particulièrement à même de profiter des corrélations spatiales entre blocs de l'image pour disperser le bruit du document hôte. Ils réalisent un entrelacement aléatoire contrôlé, qui permet d'éviter les cas les plus défavorables. Leurs performances en image sont meilleures que celles de DS dans le domaine spatial et dans la version de base. Avec l'ajout d'autres éléments dans la chaîne de tatouage (pré-blanchiment, pré-annulation des interférences, quantification) ou le passage dans un domaine transformé (DCT par blocs), ces singularités sont moins importantes. Néanmoins, les filtres LPTV permettent notamment d'atteindre une plus grande capacité pour le LISS. D'autre part, nous avons proposé une technique appelée ZI-LPTV selon une démarche originale. Elle effectue une modulation visant à annuler spectralement les interférences de l'image, qui divise la variance du bruit de l'hôte jusqu'à un facteur 100. Il s'agit d'une sorte de "tatouage à information sur le spectre de l'hôte". La problématique d'une modulation en fonction des contraintes spectrales de l'hôte est souvent mise de côté au profit d'une adaptation spatiale adaptative à l'hôte, plus performante mais avec laquelle ZI-LPTV peut être combinée. Les filtres LPTV se placent dans le cadre classique du tatouage par modulation du message puis par transmission dans un canal additif bruité. Néanmoins, ils constituent un outil théorique utile pour combiner étalement de spectre, entrelacement temporel et contraintes spectrales

sur le tatouage ou le décodeur.

La deuxième contribution principale de ce travail a été de souligner les liens entre tatouage et interpolation, *via* la contrainte d'imperceptibilité. Les attaques géométriques introduisent un bruit d'interpolation non négligeable. Nous avons souligné le fait que ce bruit d'interpolation subsiste après n'importe quelle technique de resynchronisation. Inversement, nous avons cherché à faire jouer à l'interpolation un rôle positif dans des algorithmes de tatouage. La première technique consiste à construire des masques psychovisuels à partir de l'erreur d'interpolation. Nous avons montré la cohérence de cette approche avec les masques perceptuels spatiaux classiques, ainsi que ses bonnes propriétés perceptuelles. Dans un second temps, nous avons proposé un algorithme de tatouage spécifiquement construit autour de l'interpolation. La démarche est originale, et l'algorithme bénéficie également de bonnes propriétés perceptuelles. Les règles d'insertion et de décodage peuvent être reliées aux techniques de catégorisation aléatoire robustes aux transformations valumétriques. L'algorithme bénéficie donc de bonnes performances au décodage, et surtout de rejet des interférences de l'hôte dans le cadre d'un tatouage bas débit. Des stratégies d'insertion informée ont également été proposées. La sécurité de la technique a été analysée, et des algorithmes pratiques d'attaque sur la sécurité spécifiques à la technique proposée ont été introduits. Au bilan, l'algorithme proposé est moins robuste à l'attaque AWGN, et moins sûr que les techniques de catégorisation aléatoires classiques. Cependant, il possède de bonnes propriétés perceptuelles. De plus, son application à l'image est plus robuste à de nombreuses attaques, et en particulier aux transformations valumétriques.

Les algorithmes proposés ne prennent toute leur mesure que dans le domaine spatial (robustesse pour les filtres LPTV, imperceptibilité pour l'interpolation), ce qui constitue une limitation commune à ces travaux. Les applications qui interdisent l'utilisation d'un domaine transformé, par exemple pour des raisons de complexité calculatoire, sont assez peu nombreuses. Il est donc courant en tatouage de privilégier les domaines transformés comme la transformée en ondelettes, pour des raisons d'imperceptibilité et de robustesse à certaines attaques. Dans le but d'étendre le champ d'application des techniques proposées, il nous semble donc important dans le futur d'identifier les propriétés particulières des documents dans le domaine transformé (non-stationnarité, corrélation, propriétés psychovisuelles...) qu'elles seraient en mesure d'exploiter.

Cette étude a mis en avant deux problématiques dont l'exploitation peut servir de prolongation à nos travaux. D'une part, les propriétés de corrélation du document hôte peuvent bénéficier au tatouage numérique. Notamment, nous pensons que la cyclostationnarité des signaux (introduite notamment en sortie d'un filtre LPTV) mérite d'être étudiée au bénéfice du tatouage (par exemple, pour améliorer la détection). La cyclostationnarité n'a pour l'instant été étudiée dans le tatouage que dans une application au tatouage asymétrique, par insertion d'un tatouage cyclostationnaire [dCTG02]. D'autre part, de nombreuses variantes peuvent être apportées aux techniques de tatouage fondé sur l'interpolation que nous avons proposées, car nous nous sommes surtout attaché à la généricité de nos propositions. Il serait intéressant de soumettre diverses techniques et grilles d'interpolation à des études perceptuelles subjectives, afin d'identifier des cas particuliers de masques perceptuels ou de variantes de W-interp offrant les mêmes garanties perceptuelles que les masques classiques utilisés en tatouage. Les techniques d'interpolation préservant les contours nous semblent constituer de bons candidats.

Les travaux proposés dans cette thèse ne se situent pas dans la droite lignée des problématiques les plus étudiées actuellement en tatouage numérique. Nous espérons de ce fait avoir apporté un regard neuf et original sur le sujet, tout en ayant situé nos propositions de manière objective par rapport à l'état de l'art.

Annexe A

Etude expérimentale de la robustesse des filtres LPTV

Sommaire

Dans cette annexe, on compare la robustesse des techniques de tatouage fondées sur les filtres LPTV et des techniques classiques DS, DS+W et LISS. Nous nous concentrons sur les techniques d'étalement de spectre et ses améliorations. Le chapitre 2 contient en outre une étude des techniques LPTV-SCS et du décodeur optimal dans le domaine de la DCT, et le chapitre 4 contient des comparaisons entre DS, DS+W, LISS et des techniques de catégorisation aléatoire (SCS, RDM, ST-SCS).

A.1 Etude de la robustesse des PCC : application à l'image

A.1.1 Tatouage non informé

Cadre des simulations

Ce paragraphe présente une comparaison des méthodes DS, 1D-PCC et 2D-PCC au travers de simulations. Dans les algorithmes comparés, seules les étapes Sp et Sp^{-1} diffèrent afin d'évaluer objectivement les performances des PCC. Lorsque c'est possible, nous superposons les performances théoriques des algorithmes avec les performances obtenues par simulations sur des images naturelles. La non-stationnarité et non-gaussianité de l'image entraînent des différences entre les résultats expérimentaux et théoriques. Si l'on fait varier les images $\mathbf{x}$ à tatouage $\mathbf{w}$ fixé, les résultats sont très différents d'un tatouage à l'autre (même avec 20000 images, cf. [CMB02], p. 170). Il

faut donc plutôt faire varier **w** à **x** donnée. Dans les simulations de ce document, on fera varier à la fois **w** (100 itérations par image) et **x**. On verra que la non-stationnarité a plus d'impact sur les PCC que sur DS. En effet, la "mise en forme aléatoire" de DS a pour effet de décorréler totalement les échantillons au sein d'un support $\mathcal{S}_l$, ce qui n'est pas le cas des PCC (même aléatoires) avec mise en forme répétition. Les simulations fournissent une performance moyenne sur un ensemble d'images test composé de Lena, Babouin, Bateaux, Pentagone et Poivrons (cf. *fig.* A.1).

FIG. A.1 – Images utilisées lors des simulations

Paramètres des simulations : le TEB est estimé en fonction de DWR, WNR, R (débit du message) ou J. Sauf indication contraire, les valeurs des paramètres utilisées sont : $L = 100$ bits (longueur du message), $N = 2^{18}$ pixels (taille de l'image) et $J = 1$ (tatouage simple). $T_{1D} = 2^{12}$ (période des 1D-PCC) et $T_{2D} = 2^6$ (période des 2D-PCC) sont un compromis entre l'imperceptibilité, les performances du décodage et le coût calculatoire. Pour les simulations utilisant une valeur fixe du DWR, on utilise DWR=20 dB (domaine spatial) et DWR=25 dB (DCT) afin d'avoir un TEB de l'ordre de 10^{-2} pour la chaîne de tatouage simple utilisée dans cette section. Rappelons que l'on considère généralement en tatouage d'images qu'un tatouage est imperceptible pour PSNR> 36 dB. Sur les images utilisées dans ces simulations, DWR=20 dB correspondra à PSNR=35,5 dB en moyenne.

Niveau de confiance de l'estimation du TEB : Afin d'avoir une estimation précise du TEB, les messages sont générés aléatoirement jusqu'à ce qu'au moins 100 bits erronés aient été observés. Cette règle empirique utilisée en télécommunications permet d'obtenir une confiance de 95% sur le fait que l'erreur d'estimation est inférieure à 20%. En effet, supposons que les instants d'erreurs forment un processus de Poisson de paramètre λ=TEB$_{\text{idéal}}$ (nombre moyen d'erreurs par unité de temps, *i.e.* par bit transmis). Alors la probabilité d'observer n erreurs sur un intervalle de durée τ est $\mathrm{p}(\tau, n) = \frac{(\lambda\tau)^n}{n!} e^{-\lambda\tau}$. Si l'on arrête les simulations après avoir observé 100 erreurs, en moyenne $\lambda\tau = 100$. La probabilité d'observer entre 80 et 120 erreurs est donc égale à $\sum_{n=80}^{120} \frac{(100)^n}{n!} e^{-100} = 0,9599$.

Choix des permutations optimales : toutes les permutations n'ont pas les mêmes performances sur une image non stationnaire (ce qui se traduit, à image variable, par une variation de la "mise en forme PCC" de chaque bit et de sa dispersion). Le TEB dépend de $\{f_j\}_{j\in 1\ldots J}$ car chaque permutation q_j influe d'une part sur l'étalement de **x**, et d'autre part sur les MAI.Il est possible de choisir des permutations optimales, car le nombre de permutations est fini pour T donnée. Cependant, pour les valeurs de la

période T (de l'ordre de 2^{12}) utilisées, le coût calculatoire de telles solutions, même en utilisant des algorithmes heuristiques, serait trop important. De plus, la contrainte de sécurité impose un large choix de q_j à l'insertion. Nous choisissons donc dans la suite de générer les permutations q_j aléatoirement, et de ne pas prendre en compte les MAI, ce qui est confirmé par la quasi-orthogonalité observée en pratique.

Robustesse au bruit de l'image

La *fig.* A.2 montre les performances au décodage en fonction de DWR. En-dessous de 35 dB, le tatouage est réputé perceptible. La redondance et le décodage moyenné fournissent une robustesse au bruit de l'image similaire pour DS, 1D-PCC et 2D-PCC.

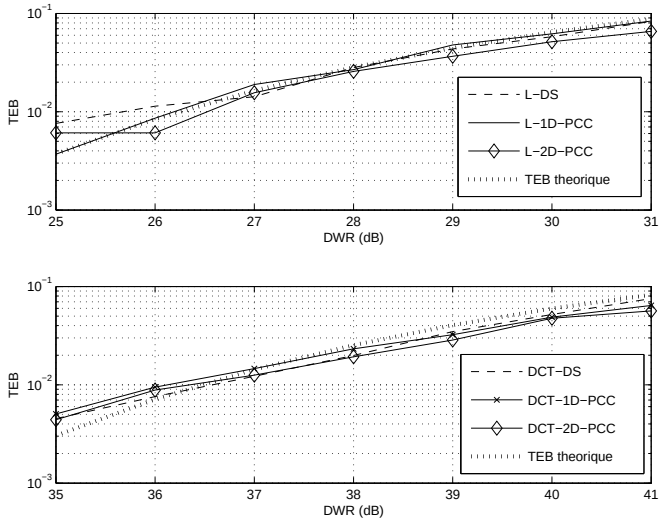

FIG. A.2 – Performance au décodage par rapport à DWR

Choix de la période des PCC pour 1D-PCC : lorsque la période T_{1D} d'une permutation augmente, ses propriétés d'orthogonalité et d'étalement d'un message binaire s'améliorent (cf. *fig.* A.3 dans le domaine spatial). Dans le domaine de la DCT, la variation du TEB est moins importante (*fig.* A.4). Dans les deux cas, $T_{1D} = 2^{12}$ offre un bon compromis entre performance et coût calculatoire.

Choix de la période des PCC pour 2D-PCC : empiriquement, $T_{2D} = 64$ semble un bon compromis entre performance au décodage et imperceptibilité pour l'ensemble des images test pour 2D-PCC-L (cf. *fig.* A.3). L'imperceptibilité impose que T soit supérieure à une période minimale. On utilisera par la suite $T_{2D} = 64$ comme période par défaut pour 2D-PCC-DCT. Dans le domaine de la DCT, les performances sont moins dépendantes de T_{2D} (cf. *fig.* A.4). En effet, après application du masque, les coefficients des moyennes fréquences de la DCT se rapprochent d'un bruit.

Influence de la longueur du message : la charge utile L limite la redondance des trois algorithmes. Cependant, elle joue un rôle particulier pour les permutations 2D. Les performances de 2D-PCC évoluent en fonction du plus petit commun multiple de N_1 et L (cf. *fig.* A.5). En effet, la répétition du message dans l'équation (2.1) introduit une décorrélation supplémentaire dans l'équation (2.3) entre les points de l'image

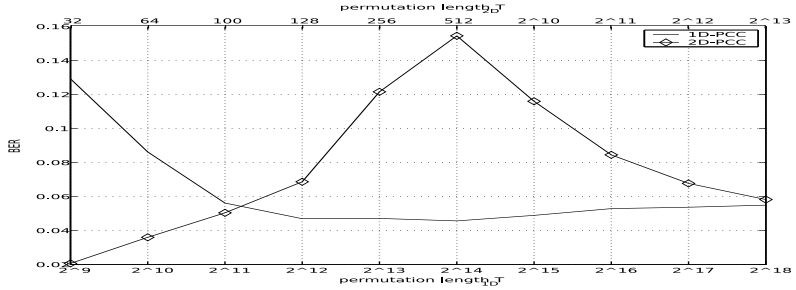

FIG. A.3 – Performance de PCC au décodage par rapport à T (Spatial), $J = 8$

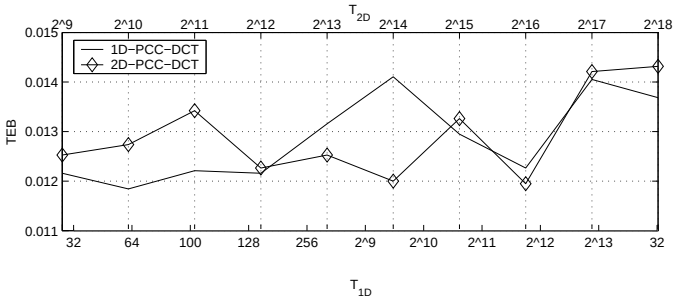

FIG. A.4 – Performance de PCC au décodage par rapport à T (DCT), $J = 8$

moyennés à la réception, sauf si N_1 est multiple de L (auquel cas, les lignes du message répété seront identiques avant étalement). Dans le cas 1D, T_{1D} est suffisamment grande par rapport à N_1 pour que cela ne joue aucun rôle. Par contre, dans le cas 2D, cela peut rendre la permutation sur les lignes inutile (les lignes étant identiques entre elles à l'issue de la permutation sur les colonnes). Cette influence n'est cependant pas très gênante car on peut facilement ajouter quelques bits à un message afin d'optimiser ppcm(N_1, L). Dans la suite, on prendra par défaut $L = 100$. Les performances dépendent également du lien entre L et T : pour certaines valeurs de L, les performances sont même indépendantes de T_{1D} ou T_{2D}.

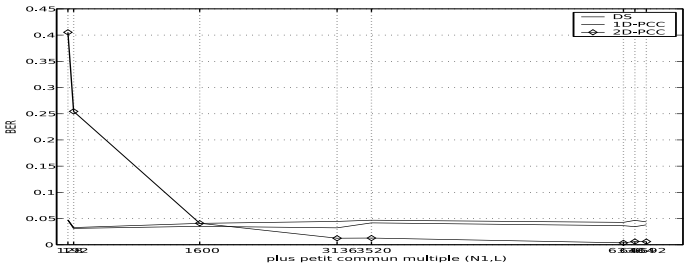

FIG. A.5 – Performance par rapport à ppcm(N_1, L) (domaine spatial)

Robustesse au bruit additif gaussien

La *fig.* A.6 montre que les performances à la détection des trois algorithmes est inchangée après l'application d'un bruit additif, dans le domaine spatial comme dans celui de la DCT. La région de droite (WNR> 0) correspond à une attaque imperceptible

tible. Le TEB y est constant, ce qui montre que les algorithmes sont affectés par l'image hôte mais pas par l'attaque. Dans la région de droite, le TEB augmente mais l'attaque est perceptible.

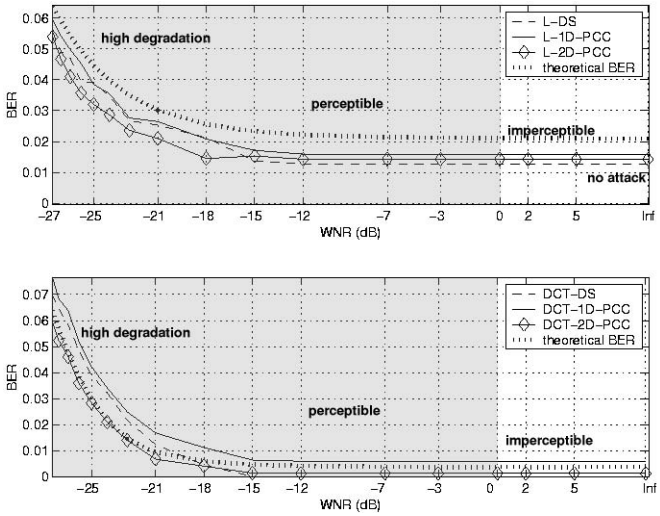

FIG. A.6 – Robustesse à une attaque de type bruit additif

Influence du débit du message et du tatouage multiple

Soit $R = 1/P$ le débit. La performance au décodage de DS, 1D-PCC et 2D-PCC est tout d'abord estimée en fonction de la redondance P introduite dans le message. Ces performances sont supposées augmenter avec P. Par exemple, pour $P = 1$ (pas de redondance), un mauvais TEB (TEB=10^{-2}) n'est atteint qu'au prix d'une détérioration de l'image (on doit avoir DWR=-5 dB). La *fig.* A.7 montre que DS et PCC se comportent de manière similaire lorsque R augmente, avec une légère supériorité de 2D-PCC dans le domaine spatial (rappelons cependant que les performances de 2D-PCC dépendent fortement de $L = N/P$).

Le tatouage multiple offre une autre façon de transmettre un nombre donné de bits. Les *fig.* A.8 et A.9 montrent les performances par rapport à J dans le cas du tatouage multiple. Les tests ont été effectués avec $J = 2$ à 14 messages. Le débit multiple défini par $R_J = JL/N$ augmente proportionnellement. Les algorithmes ont des performances similaires lorsque J est grand (cf. *fig.* A.9). Les performances dépendent uniquement de R_J : le partage de la puissance de $\mathbf{m}$ entre plusieurs utilisateurs orthogonaux, compte tenu des interférences multi-utilisateurs, fournit des résultats équivalents au partage spatial de l'image entre les bits d'un seul message, où P diminue.

Robustesse à des attaques sophistiquées

Débruitage (filtrage de Wiener) : les simulations montrent une grande dégradation des performances (cf. *fig.* A.10). PCC-DCT est un peu plus robuste au filtrage de Wiener que DS-DCT.

Compression JPEG : $\mathbf{y}$ est compressée au format JPEG avec un taux de compression donné. Bien que la déformation visuelle soit très légère, l'attaque est très efficace sur les trois algorithmes dans le domaine spatial (*fig.*A.11). Dans le domaine de la DCT

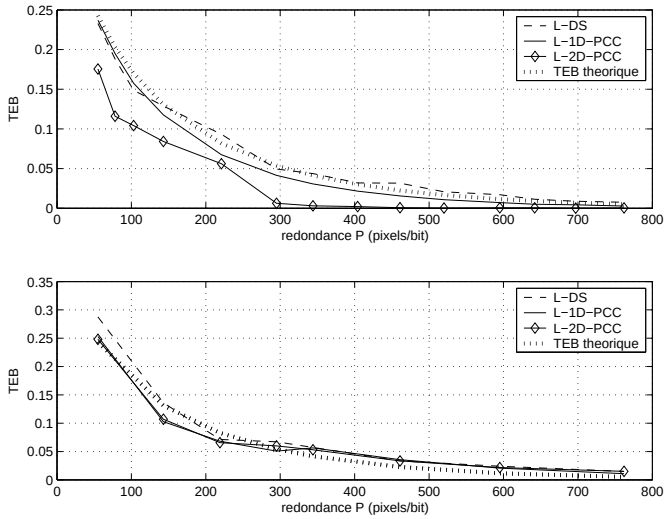

FIG. A.7 – Performance au décodage par rapport à P

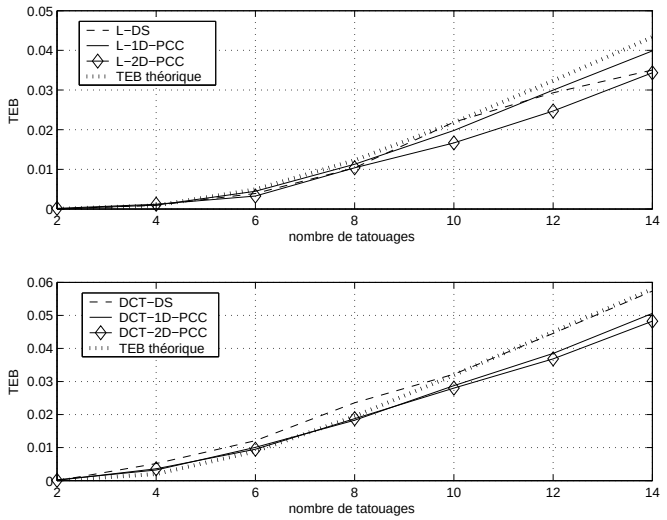

FIG. A.8 – Performance au décodage par rapport à J

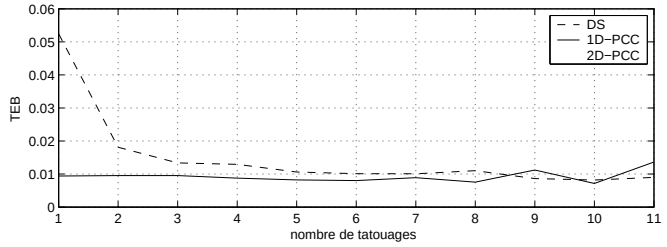

FIG. A.9 – Performance au décodage par rapport à J, avec R_J constant

(cf. *fig.* A.11), la robustesse est logiquement beaucoup plus grande, car les moyennes fréquences sont peu affectées par la quantification avec perte.

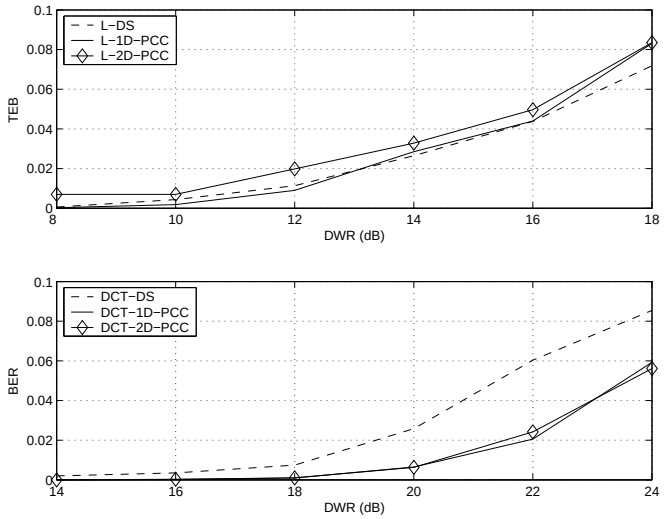

FIG. A.10 – Robustesse au filtrage de Wiener

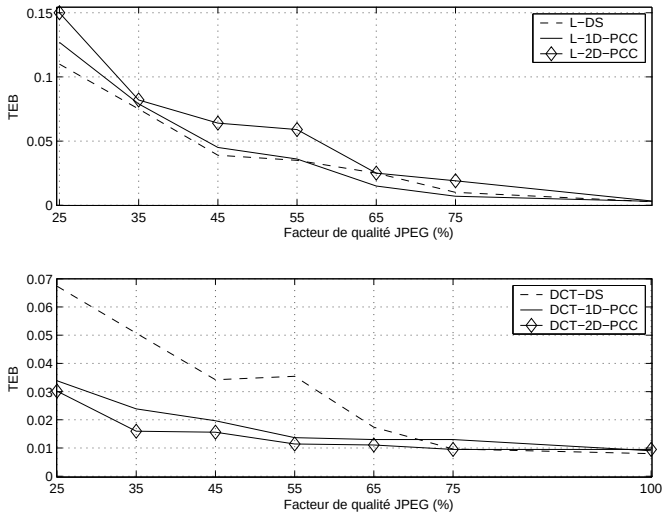

FIG. A.11 – Robustesse à la compression JPEG en fonction du facteur de qualité

A.2 Etude de la robustesse des filtres LPTV : application à l'image

A.2.1 Tatouage non informé

Interférences de l'hôte

Influence de la période : les performances de LL-LPTV, mod-LPTV et mod-NRZ-LPTV s'améliorent lorsque T augmente, mais convergent vers un palier (cf. *fig.* A.12). Les performances de LL-NRZ-LPTV sont indépendantes de T. En pratique, le coût calculatoire est trop important si $T > 1024$. Dans les simulations qui suivent, on prendra $T = 128$, qui est un bon compromis entre performance et temps de calcul. Cependant, dans des applications pratiques sans contraintes temps réel ou avec des implantations optimisées, il est parfaitement envisageable d'utiliser une période T importante, afin

d'augmenter la sécurité de l'algorithme.

Puissance d'insertion : LL-NRZ-LPTV et mod-NRZ-LPTV sont moins robustes au bruit de l'image que les autres techniques du fait de la mise en forme NRZ (cf. *fig.* A.13). mod-LPTV offre de meilleures performances que DS, et LL-LPTV de meilleures performances que mod-NRZ. Quant à ZI-LPTV, il présente les meilleures performances grâce à l'élimination des basses fréquences de l'image.

Débit : LL-LPTV et mod-LPTV sont très sensibles à L à cause de la mise en forme répétition (cf. *fig.* A.14). Comme $T = 128$ est une puissance de 2, leurs performances sont notamment très mauvaises pour $L = 128, 256, 512$, *i.e.* pour L multiple de T. Cependant, LL-LPTV offre de meilleures performances que mod-LPTV, et mod-LPTV de meilleures performances que DS. Les performances de ZI-LPTV, qui est peu sensible au choix de L grâce à la mise en forme NRZ, sont de loin les meilleures.

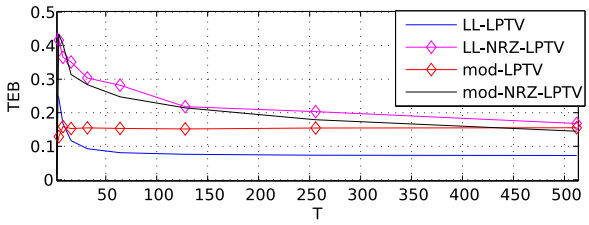

FIG. A.12 – Influence de T, LL-LPTV, DWR=28dB, $L = 512$

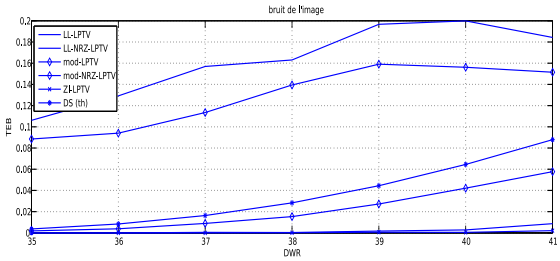

FIG. A.13 – Influence de la puissance d'insertion, $L = 100$

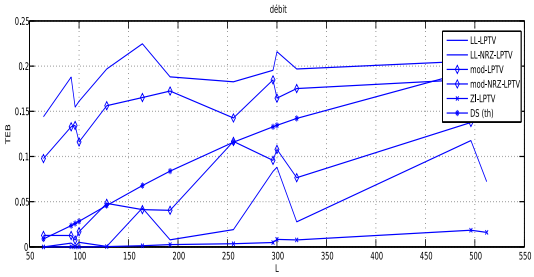

FIG. A.14 – Influence de L, DWR=28 dB

Attaques sur la robustesse

Compression JPEG : les trois techniques utilisant les filtres LPTV et le parcours de Peano (PCC, LL-LPTV et mod-LPTV) fournissent des performances à la compres-

sion JPEG supérieures à celles de DS (cf. *fig.* A.15). ZI-LPTV fournit une excellente robustesse à la compression JPEG.

Débruitage : LL-LPTV est toujours très sensible à L (cf. *fig.* A.16). ZI-LPTV offre une excellente robustesse au débruitage. Les autres techniques, y compris DS, sont fragiles au débruitage.

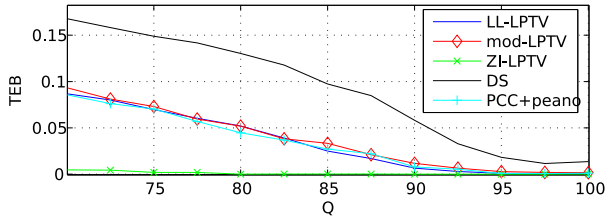

FIG. A.15 – Robustesse à la compression JPEG, DWR=25 dB, $L = 100$

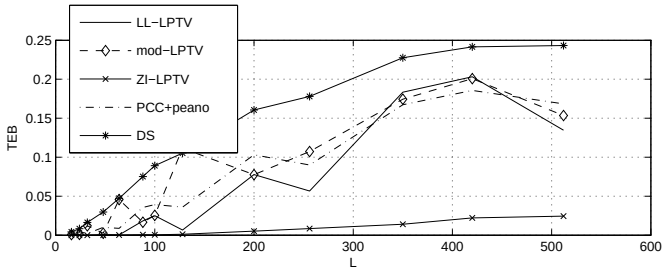

FIG. A.16 – Robustesse au débruitage, DWR=28 dB

Insertion dans un domaine transformé

Comme pour les algorithmes DS et PCC, l'insertion peut avoir lieu dans un domaine transformé approprié. Les performances de l'insertion dans le domaine de la DCT par blocs sont présentées sur la *fig.* 2.49. La robustesse au bruit de l'image est similaire dans le domaine transformé. Comme pour les PCC, l'insertion dans un domaine transformé bénéficie par contre à l'imperceptibilité et à la robustesse à des attaques telles que la compression JPEG. Notons que le domaine de la DCT par blocs possède des propriétés statistiques particulières. Par exemple, un parcours de Peano de taille 8 parcourt les 64 pixels d'un bloc, dont la corrélation pourrait être exploitée par les filtres LPTV.

A.2.2 Pré-blanchiment

PCC

Le préfiltrage de Wiener apporte une nette amélioration des performances des PCC (cf. *fig.* A.17 à A.21, à comparer avec les *fig.* A.6 à A.11). Pour un TEB donné, l'amélioration du DWR est de l'ordre de 7 dB si $L = 100$. Nous étudions les performances pour une valeur DWR correspondant à l'imperceptibilité théorique DWR_{imp} spécifique à chaque image. Les nouvelles valeurs par défaut choisies dans les simulations qui

suivent sont DWR=DWR$_{\text{imp}}$, $P = 100$ et donc $L = N/P = 2621$. Pour ces valeurs, les performances sont peu dépendantes de T, même pour les 2D-PCC dans le domaine spatial (cf. *fig.* A.21).

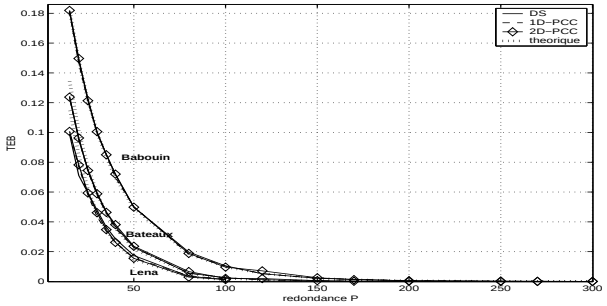

FIG. A.17 – Performance au décodage par rapport à la redondance P (préfiltrage de Wiener au décodage) avec DWR=DWR$_{\text{imp}}$ pour chaque image

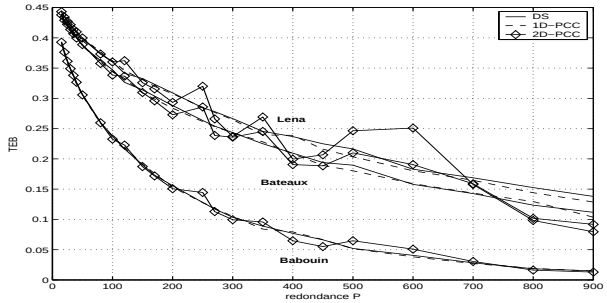

FIG. A.18 – Robustesse à l'attaque de débruitage (préfiltrage de Wiener au décodage)

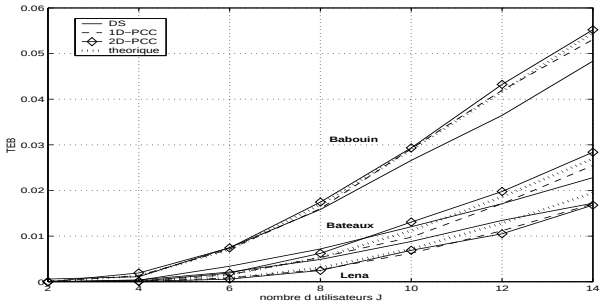

FIG. A.19 – Performance au décodage par rapport à J (préfiltrage de Wiener au décodage) (P=700)

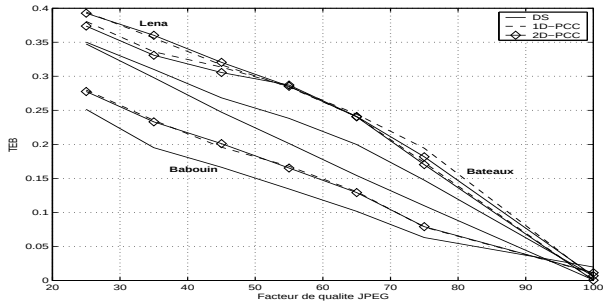

FIG. A.20 – Robustesse à la compression JPEG en fonction du facteur de qualité (préfiltrage de Wiener au décodage, $P = 100$)

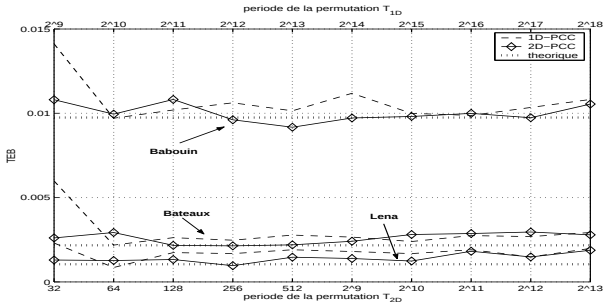

FIG. A.21 – Performance de PCC au décodage par rapport à T (préfiltrage de Wiener au décodage)

Filtres LPTV

Puissance d'insertion : mod-LPTV+W bénéficie peu du filtrage de Wiener, avec des performances (hors figure), de l'ordre de 10^{-2} dB (cf. *fig.* A.22). LL-NRZ-LPTV+W et mod-NRZ-LPTV+W bénéficient grandement du préfiltrage de Wiener qui élimine une grande partie des interférences dues à la non-stationnarité de l'image. Les simulations, peu précises car le TEB est faible, montrent des performances similaires pour DS+W, LL-LPTV+W et ZI-LPTV+W.

Débit : mod-LPTV+W et LL-LPTV+W sont toujours très sensibles au choix de L (cf. *fig.* A.23). LL-NRZ-LPTV+W et mod-NRZ-LPTV+W sont peu sensibles au choix de L, et dans certains cas elles offrent de meilleures performances que LL-LPTV+W. ZI-LPTV+W offre les meilleures performances mais l'amélioration par rapport à mod-NRZ-LPTV est beaucoup plus faible que sans préfiltrage au décodage, car le préfiltrage élimine une grande partie des composantes de l'image dans les basses fréquences.

Compression JPEG : les techniques PCC, LL-LPTV et mod-LPTV bénéficient d'une façon similaire à DS+W du préfiltrage de Wiener au décodage (cf. *fig.* A.24).

Débruitage : DS+W est la technique la plus robuste au débruitage : DS bénéficie plus du préfiltrage de Wiener que les autres techniques (cf. *fig.* A.25). mod-NRZ-LPTV+W et LL-NRZ-LPTV+W offrent cette fois-ci une meilleure robustesse que leurs homologues avec répétition.

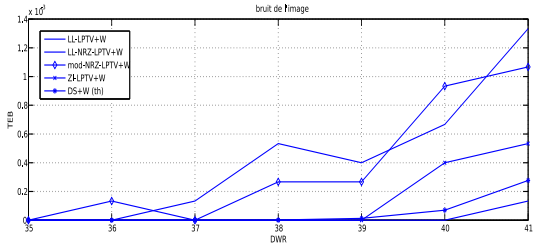

FIG. A.22 – Préfiltrage de Wiener : influence de la puissance d'insertion, $L = 100$

A.2.3 Étalement de spectre amélioré

PCC

Les *fig.* A.26 et A.27 montrent la supériorité de IPCC+peano sur LISS : on profite toujours de la propriété $\sigma^2_{\mathbf{x},\text{PCC, Peano}} < \sigma^2_{\mathbf{x}}$ (cf. partie 2.2)).

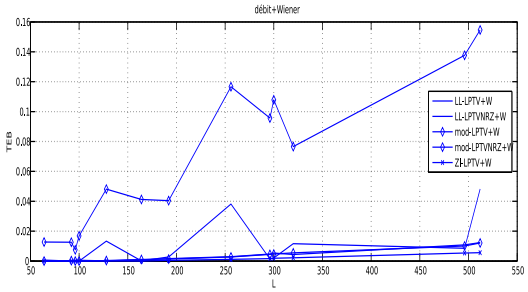

FIG. A.23 – Préfiltrage de Wiener : influence de L, DWR=28 dB

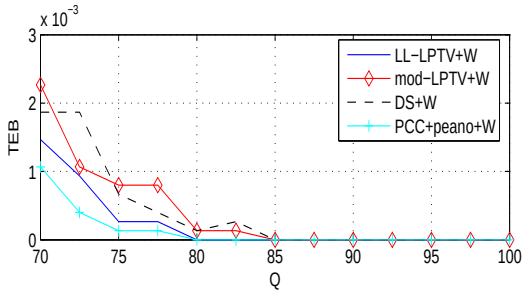

FIG. A.24 – Préfiltrage de Wiener : robustesse à la compression JPEG, DWR=25 dB, $L = 100$

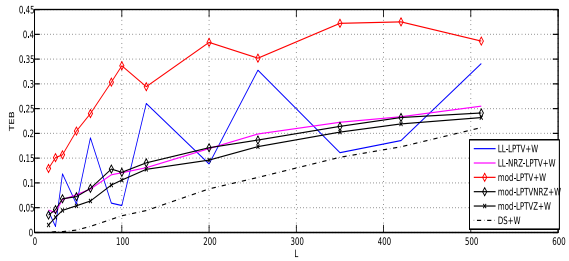

FIG. A.25 – Préfiltrage de Wiener : robustesse au débruitage, DWR=28 dB

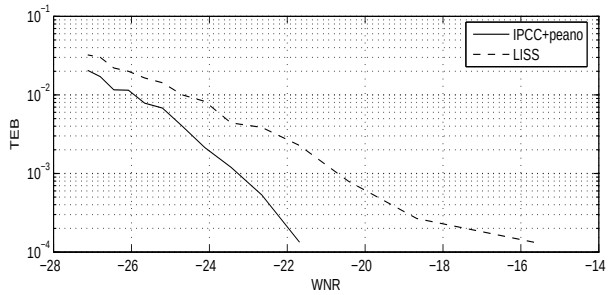

FIG. A.26 – IPCC+peano : performance en présence de bruit AWGN, DWR=28, $L = 100$

Filtres LPTV

Puissance d'insertion : les trois techniques LPTV améliorées (LL-ILPTV, mod-ILPTV et IPCC+peano) bénéficient du préfiltrage à l'insertion. De plus, le parcours de Peano combiné à un étalement LPTV réduit le bruit d'un document image. Leurs performances sont donc mêmes meilleures que celles de LISS lorsque le DWR est trop élevé pour qu'on ait une annulation des interférences de l'image. ZI-ILPTV, qui élimine

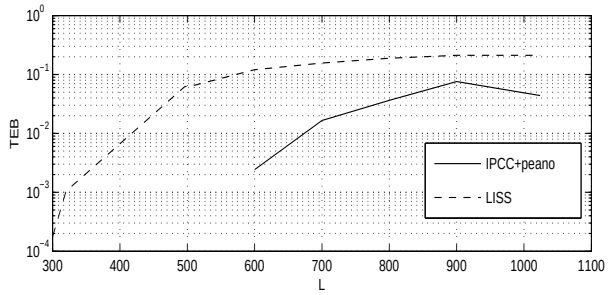

FIG. A.27 – IPCC+peano : performance en fonction de L, DWR=28

une grande partie des interférences de l'hôte, présente d'excellentes performances.

Bruit AWGN : LL-ILPTV et IPCC+Peano sont les techniques les plus robustes à un fort bruit AWGN.

ZI-LPTV : la *fig.* A.30 montre que ZI-LPTV bénéficie du préfiltrage de Wiener et/ou de l'insertion informée, mais dans une moindre mesure que les autres filtres LPTV.

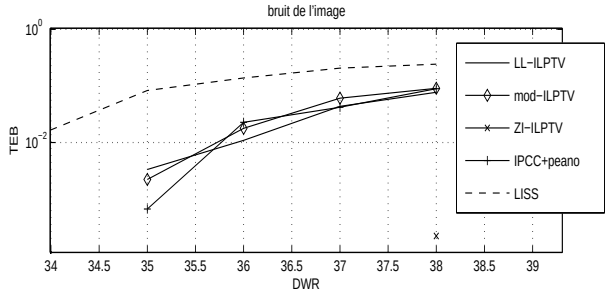

FIG. A.28 – LPTV améliorés : influence du DWR, L=100

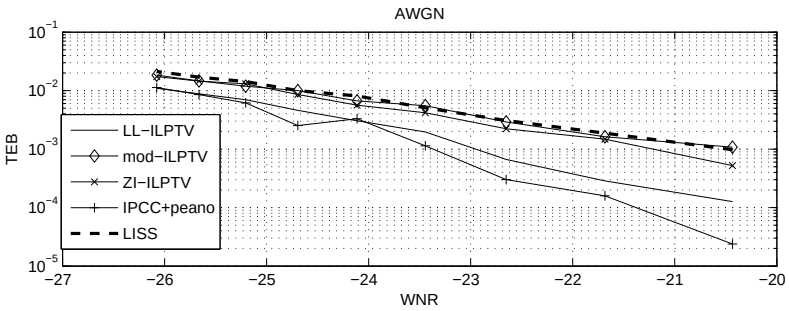

FIG. A.29 – LPTV améliorés : robustesse au bruit AWGN, DWR=28dB, L=100

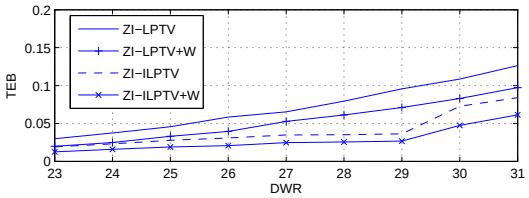

FIG. A.30 – Intérêt du préfiltrage de Wiener et de l'insertion informée pour ZI-LPTV, $L = 2048$

A.3 Robustesse à un bruit d'interpolation : application à l'image

Filtres LPTV et changement d'échelle : la *fig.* A.31 présente la robustesse face à un changement d'échelle suivi d'une resynchronisation, en fonction du DNR introduit par l'attaque. On y retrouve la supériorité de ZI-LPTV sur les autres techniques, et de mod-LPTV, LL-LPTV et PCC+peano sur DS.

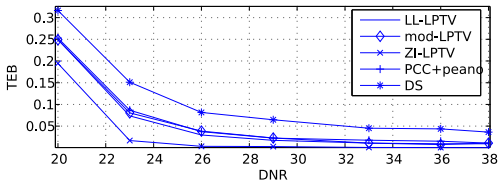

FIG. A.31 – Robustesse au bruit d'interpolation, DWR=25 dB, $L = 100$

Annexe B

Annexe sur les filtres LPTV

Sommaire

Dans cette annexe, nous détaillons des propriétés particulières des filtres LPTV. L'étude des performances théoriques des PCC est approfondie. Nous montrons ensuite que les PCC sont applicables au tatouage multiplicatif. Puis nous détaillons les algorithmes pratiques d'attaque sur la sécurité évoqués dans la partie 2.6.

B.1 Performances théoriques des PCC

B.1.1 Capacités d'étalement des PCC

Pour T suffisamment grand, le spectre de V s'approche de celui d'un bruit blanc [LR02]. En effet, si le spectre de puissance $S_X(\omega)$ de X est défini par :

$$K_X(n) = E[x_m x^*_{m-n}] = \int_{-\pi}^{\pi} e^{in\omega} S_X(\omega) d\omega \quad \forall m \in \mathbb{X} \ ,$$

on a :

$$\left\{ \begin{array}{lcl} K_V(n) & = & K_X(n) \text{ si } \underline{n} = 0 \\ K_V(n) & = & \frac{1}{T} \int_{-\pi}^{\pi} e^{i\omega T \overline{n}} [\underline{n} e^{iT\omega} + (T - \underline{n}) a_T(\omega)] S_X(\omega) d\omega \text{ sinon} \ , \end{array} \right.$$

$$\text{où } a_T(\omega) = \frac{1}{T(T-1)} \left[\left(\frac{\sin T\omega/2}{\sin \omega/2} \right)^2 - T \right]$$

(voir [LR02] pour le détail des calculs). En supposant S_X suffisamment régulière, on a : $\lim_{n\to\infty} K_X(n) = 0$ [1]. Alors [2]

$$\lim_{T\to\infty} K_V(n) = \begin{cases} K_X(0), n = 0 \\ 0, n \neq 0 \end{cases} .$$

D'après les propriétés classiques des fonctions caractéristiques, la limite précédente implique :

$$\lim_{T\to\infty} \int_a^b S_V(\omega)d\omega = \frac{K_X(0)}{2\pi}(b-a), -\pi \leq a < b \leq \pi$$

et le spectre de V est donc étalé.

B.1.2 Détecteurs multi-symboles pour estimation-détection conjointes

Lorsque le message comporte plusieurs bits, on ne peut pas travailler avec $\sum_{l=1}^{L} d_l^j$ car les valeurs risquent de s'annuler. Plusieurs stratégies sont alors possibles. La première stratégie consiste à détecter chaque symbole indépendamment par corrélation linéaire (mais pas normalisée). Si un tatouage est détecté sur le support d'au moins 1 bit, on détecte le tatouage [CMB02].

Le **détecteur optimal** consiste en un test d'hypothèses composite [Ver98]. Pour L bits d'information, soient $\{H_i, i = 1 \dots 2^L\}$ les hypothèses correspondant à la présence des 2^L messages $\mathbf{m}_i^j$ possibles. On travaille alors sur les hypothèses composites $\bar{H}_0$ et $\bar{H}_1$:

$$\bar{H}_1 : \begin{cases} H_1 \\ \vdots \\ H_{2^L} \end{cases} \text{ (présence d'un message } \mathbf{m}_i^j \text{ donné)}$$

$$\bar{H}_0 : H_0 \text{ (absence de message)}$$

La règle de décision optimale est alors :

$$\exp\left(\frac{1}{\sigma_{\mathbf{x}}^2} < \psi P\mathbf{d}^j, 0 >\right) = 0 \lessgtr \sum_{i=1}^{2^L} \exp\left(\frac{1}{\sigma_{\mathbf{x}}^2} < \psi P\mathbf{d}^j, \mathbf{m}_i^j >\right)$$

En pratique, pour $L > 20$, le coût calculatoire de ce détecteur est trop élevé.

Nous proposons dans ce paragraphe d'utiliser un **détecteur d'énergie**, plus simple, qui utilise comme observations : $r = \frac{P}{\sigma_{\mathbf{x}}^2} \sum_{l=1}^{L} |d_l^j|^2$, avec comme pour le détecteur binaire :

$$\bar{H}_1 : d_l^j \sim \mathcal{N}(\pm\psi, \frac{\sigma_{\mathbf{x}_1}^2}{P})$$

$$\bar{H}_0 : d_l^j \sim \mathcal{N}(0, \frac{\sigma_{\mathbf{x}_0}^2}{P})$$

[1] Lemme de Riemann-Lebesgue pour les séries de Fourier : si une fonction f est intégrable sur $[a, b]$ alors :

$$\lim_{n\to\infty} \int_a^b f(t)e^{int}\mathrm{d}t$$

[2] si $T \to \infty$ et $\underline{n} = 0$, alors $n \neq 0$, $n = \overline{n}T \to \infty$

Alors sous l'hypothèse $\bar{H}_0$, $\mathbf{T} \sim \chi^2_L$ (avec E[$R|H_0$]=L et Var[$R|H_0$]=$2L$). Sous l'hypothèse $\bar{H}_1$, on peut calculer [3] E[$R|H_1$]=$L(1+\frac{P\psi^2}{\sigma^2_{\mathbf{x}}})$ et Var[$R|H_1$]=$2L(1+2\frac{P\psi^2}{\sigma^2_{\mathbf{x}}})$. Pour L suffisamment grand (à partir de $L > 20$), on peut utiliser le Théorème de la Limite Centrale pour remplacer le χ^2_L par une gaussienne. Le **détecteur d'énergie simplifié** est alors :

$$\bar{H}_1 : R \sim \mathcal{N}(L(1+\frac{P\psi^2}{\sigma^2_{\mathbf{x}}}), 2L(1+2\frac{P\psi^2}{\sigma^2_{\mathbf{x}}}))$$
$$\bar{H}_0 : R \sim \mathcal{N}(L, 2L)$$

Et le test d'hypothèses correspondant permet d'obtenir :

$$P_{\text{fa}} = Q\left(\frac{\eta - L}{\sqrt{2L}}\right) \text{ et} \qquad P_d = 1 - Q\left(\frac{L(1+\frac{P\psi^2}{\sigma^2_{\mathbf{x}}}) - \eta}{\sqrt{2L(1+2\frac{P\psi^2}{\sigma^2_{\mathbf{x}}})}}\right)$$

La règle de décision est

$$\mathbf{T} = r^2 \frac{\frac{P\psi^2}{\sigma^2_{\mathbf{x}}}}{L(1+2\frac{P\psi^2}{\sigma^2_{\mathbf{x}}})} + r\frac{-\frac{P\psi^2}{\sigma^2_{\mathbf{x}}}}{(1+2\frac{P\psi^2}{\sigma^2_{\mathbf{x}}})} \lessgtr \ln\left(1+2\frac{P\psi^2}{\sigma^2_{\mathbf{x}}}\right) + \frac{L}{2}\frac{(\frac{P\psi^2}{\sigma^2_{\mathbf{x}}})^2}{(1+2\frac{P\psi^2}{\sigma^2_{\mathbf{x}}})}$$

Hernandez *et al.* [HPGRN98] proposent également d'effectuer une détection multiple, sans passer par un détecteur d'énergie et en tenant compte des probabilités de chaque bit. L'implantation de ce détecteur est cependant très complexe.

En pratique, la probabilité de fausse alarme est très faible au seuil d'imperceptibilité (DWR=26 dB), ce qui rend impossible un calcul expérimental de P_{fa}. Pour des valeurs de DWR plus élevées, on obtient les courbes COR des *fig.* B.2 et B.1. La formule théorique du détecteur d'énergie simplifiée est cohérente avec les simulations. Là encore, 2D-PCC fournit de meilleures performances que DS et 1D-PCC.

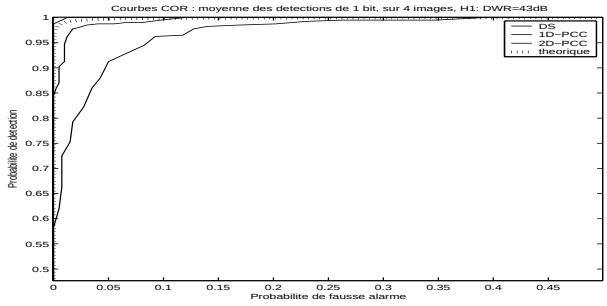

FIG. B.1 – Courbe COR : détection de 100 bits, bruit additif fort : WNR=-29 dB

[3] Var[$R|H_1$]=LVar[$|d^j_l|^2|H_1$]=L(E[$|d^j_l|^4$]-E[$|d^j_l|^2$]^2).
Si $X \sim \mathcal{N}(\mu, \sigma^2)$, on a E[$X^k$]=$\frac{\mathrm{d}^k}{\mathrm{d}t^k} M_X(t)|_{t=0}$,
où $M_X(t)$ est la fonction génératrice des moments : $M_X(t)$=E[e^{tX}]= $e^{(t\mu + t^2 * \sigma^2/2)}$.
Donc E[$|X|^2$]=$(\sigma^2+\mu^2)$ et E[$|X|^4$]=$3\sigma^4 + 6\sigma^2\mu^2 + \mu^4$.
Ici, $\mu = \frac{\sqrt{P}\psi}{\sigma_I}$ et $\sigma^2 = 1$, d'où le résultat.

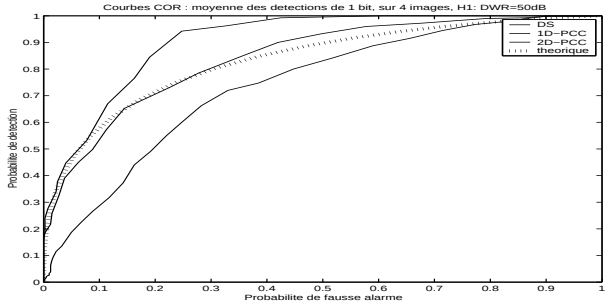

FIG. B.2 – Courbe COR : détection de 100 bits, P=2621, DWR=40 dB

B.1.3 Performance des PCC au décodage en tenant compte des MAI

Si l'on reprend la section 2.2 pour tenir compte de la corrélation entre les messages permutés,

$$\hat{d}_l^j = \psi m_l^j + \frac{1}{P}\sum_{p=1}^{P}(f_j^{-1}(\mathbf{x}+\mathbf{b}^k))(l+(p-1)L)) + \frac{\psi}{P}\sum_{p=1}^{P}\sum_{k\neq j}(f_j^{-1}\circ f_k(\mathbf{b}^k))(l+(p-1)L) \ .$$

On peut alors calculer l'expression théorique du TEB grâce à la formule de Bayes [CR03] :

$$\begin{aligned}\text{TEB}_l^j = \frac{1}{2^{(J-1)L}}\sum_{k\neq j}\sum_{M_k\in\{-1,+1\}^L} & \frac{1}{2}Q\left(\frac{\psi\sqrt{P}}{\sqrt{\sigma_{\mathbf{y}}^2+\sigma_{\mathbf{n}}^2}}(1-\frac{1}{P}\sum_{p=1}^{P}\sum_{k\neq j}(f_j^{-1}\circ f_k(\mathbf{b}^k))(l+(p-1)L))\right) \\ & + \frac{1}{2}Q\left(\frac{\psi\sqrt{P}}{\sqrt{\sigma_{\mathbf{y}}^2+\sigma_{\mathbf{n}}^2}}(1+\frac{1}{P}\sum_{p=1}^{P}\sum_{k\neq j}(f_j^{-1}\circ f_k(\mathbf{b}^k))(l+(p-1)L))\right)\end{aligned}$$

Ce TEB dépend de l'utilisateur j et du bit l, à cause des permutations. Cependant, le coût calculatoire de cette expression croît exponentiellement avec J et L. Considérons désormais $\frac{1}{P}\sum_{p=1}^{P}\sum_{k\neq j}(f_j^{-1}\circ f_k(\mathbf{b}^k))(l+(p-1)L))$ comme une variable aléatoire gaussienne de moyenne nulle et de variance $\sigma^2_{\text{MAI}}(j,l)$ Grâce au théorème de la limite centrale, on a :

$$\text{TEB}_l^j \simeq Q\left(\frac{\psi\sqrt{P}}{\sqrt{\sigma_{\mathbf{y}}^2+\sigma_{\mathbf{n}}^2+\psi^2\sigma^2_{\text{MAI}}(j,l)}}\right) \ .$$

Pour atténuer l'impact de $\sigma^2_{\text{MAI}}(j,l)$ sur le TEB, on peut utiliser un filtre décorrélateur [CR03]. Cependant, pour les valeurs de la période T (de l'ordre de 2^{12}) utilisées dans le cadre du tatouage, le coût calculatoire de ce filtre est très important. On peut également effectuer une réception optimale au sens des moindres carrés (MMSE). Le récepteur

linéaire MMSE présenté dans [CR04] peut être vu comme un compromis entre le récepteur à filtre adapté et le récepteur décorrélateur. Là encore, le coût calculatoire de ce récepteur est important.

Dans le cas du tatouage à insertion aveugle, lorsque DWR est suffisamment grand, $\psi^2\sigma^2_{\text{MAI}}(j,l) << \sigma^2_{\mathbf{x}}$. Il est donc préférable au vu des contraintes calculatoires de négliger la contribution des MAI dans le récepteur MMSE et celui-ci est alors équivalent au filtre adapté. Dans le cas du tatouage informé, les MAI doivent être prises en compte.

B.2 Adaptation des PCC au tatouage multiplicatif

Principe du schéma multiplicatif

Les schémas étudiés jusqu'à présent concernaient une insertion additive $y_k = x_k + \psi w_k$. On peut aussi adopter un schéma multiplicatif :

$$y_k = x_k + \psi w_k x_k$$

L'amplitude du tatouage est donc modulée par l'image. Un tel schéma est plus compliqué que le schéma additif (notamment au niveau de la détection car la corrélation n'est plus optimale), mais a des propriétés d'imperceptibilité intéressantes (notamment dans les domaines fréquentiels). La sécurité est également améliorée car le tatouage dépend de l'image originale, ce qui peut être une parade efficace à l'attaque de moyennage notamment. En supposant l'hôte gaussien, le calcul de la vraisemblance conduit à la statistique suffisante du décodeur optimal suivante :

$$d_l^j = \frac{1}{P}\sum_{p=1}^{P}(z(p+(l-1)L))^2 c^j(p) \ ,$$

au lieu de la corrélation $d_l^j = \frac{1}{P}\sum_{p=1}^{P} z(p+(l-1)L)c^j(p)$ dans le cas additif.

Le TEB théorique est alors :

$$TEB = Q\left(\frac{\sqrt{2P\sigma^2_{\mathbf{w}}\sigma^2_{\mathbf{x}}}}{\sigma^2_{\mathbf{x}}+\sigma^2_{\mathbf{n}}}\right) \ ,$$

au lieu de pour $TEB = Q\left(\sqrt{\frac{P\sigma^2_{\mathbf{w}}}{\sigma^2_{\mathbf{x}}+\sigma^2_{\mathbf{n}}}}\right)$ dans le cas additif. S'il n'y a pas de bruit, il y a un rapport de $\sqrt{2}$ dans l'argument, soit un gain de 3 dB pour DWR. Par contre, le schéma multiplicatif est plus sensible au bruit. Le choix des schémas dépend donc du rapport entre $\sigma^2_{\mathbf{n}}$ et $\sigma^2_{\mathbf{x}}$.

Souvent, le tatouage multiplicatif est appliqué dans le domaine fréquentiel : cette technique a été utilisée dans le domaine de la DCT par blocs [PBBC97] et de la DFT [BBRP01]. On peut trouver dans [BBR02] un autre calcul théorique des performances du tatouage multiplicatif simple (en supposant l'hôte gaussien). Comme dans [HAPG00] pour une insertion additive dans le domaine de la DCT, on peut s'appuyer sur une estimation de la densité de probabilité des coefficients de la DFT pour construire un décodeur optimal utilisant un test statistique. Plusieurs autres articles s'intéressent à la détection optimale des tatouages additifs [CH03].

PCC et schéma multiplicatif

Pour les mêmes performances théoriques, l'équivalent du schéma multiplicatif pour les PCC est :

$$\mathbf{y} = \mathbf{x} + \psi(\mathbf{x} - \mu(\mathbf{x})) \sum_{j=1}^{J} f(\mathbf{b}_j)$$

Au décodage :

$$\hat{d^j}_l = \frac{1}{P} \sum_{p=1}^{P} (f_j^{-1}((\mathbf{z} - \mu(\mathbf{z}))^2))(l + (p-1)L)$$

Le terme $(\mathbf{z} - \mu(\mathbf{z}))^2$ est nécessaire car :

- si on ne met pas le carré, la moyenne des $(x_k - \mu(\mathbf{x}))f(\mathbf{b}_j)(n)$ tend vers 0
- si on ne retranche pas la moyenne, on effectue une sorte de masque perceptuel proportionnel à l'amplitude de l'image. Celui-ci n'a pas de très bonnes propriétés perceptuelles (cf. paragraphe 1.5.3) et gêne les performances au décodage.

Bibliographie

[BBR02] M. Barni, F. Bartolini, and A. De Rosa. On the performance of multiplicative spread spectrum watermarking. *IEEE Workshop on Multimedia Signal Processing*, pages 324–327, 2002.

[BBRP01] M. Barni, F. Bartolini, A. De Rosa, and A. Piva. A new decoder for the optimum recovery of nonadditive watermarks. *Image Processing, IEEE Transactions on*, 10(5) :755–766, 2001.

[CH03] Q. Cheng and T.S. Huang. Robust optimum detection of transform domain multiplicative watermarks. *IEEE Trans. on Signal Processing*, 51(4) :906–924, 2003.

[CMB02] I.J. Cox, M.L. Miller, and J.A. Bloom. *Digital Watermarking*. Morgan Kaufmann Publisher, Inc., San Francisco, 2002.

[CR03] M. Coulon and D. Roviras. Multi-user detection for random permutation-based multiple access. *IEEE ICASSP'03, Proc.*, 4 :61–64, 2003.

[CR04] M. Coulon and D. Roviras. MMSE Joint Detection for an Asynchronous Spread-Spectrum System Based on Random Permutations. *IEEE ICASSP'04, Proc.*, 2 :17–21, 2004.

[HAPG00] J.R. Hernández, M. Amado, and F. Pérez-González. DCT-Domain Watermarking Techniques for Still images : Detector Performance analysis and a New Structure. *IEEE Trans. on Image Processing*, 9(1) :55–68, 2000.

[HPGRN98] J.R. Hernández, F. Pérez-González, J.M. Rodriguez, and G. Nieto. Performance analysis of a 2-D-multipulse amplitude modulation scheme for data hiding and watermarking of still images. *Selected Areas in Communications, IEEE Journal on*, 16(4) :510–524, 1998.

[LR02] B. Lacaze and D. Roviras. Effect of random permutations applied to random sequences and related applications. *Signal Processing*, 82 :821–831, 2002.

[PBBC97] A. Piva, M. Barni, F. Bartolini, and V. Cappellini. DCT-based watermark recovering without resorting to the uncorrupted original image. *ICIP'97, Proc.*, 1 :520–523, 1997.

[Ver98] S. Verdu. *Multiuser Detection*. Cambridge University Press, 1998.

Annexe C

Annexe sur l'interpolation

Sommaire

Dans cette annexe, nous complétons l'étude de l'algorithme W-interp. Nous justifions l'intuition d'utiliser l'interpolation au bénéfice de l'imperceptibilité d'un tatouage par une étude subjective. Nous proposons un seuil itératif de décodage empirique qui améliore les performances par rapport au seuil empirique proposé dans le chapitre 4. Nous proposons ensuite une autre technique d'insertion informée pour W-interp, utilisant des techniques d'optimisation. D'autre part, nous étudions l'impact d'une attaque sur la robustesse spécifique à W-interp. Cette étude donne l'occasion d'introduire le modèle d'image de Markov-Gauss auquel il a été fait référence dans le chapitre 4. Nous proposons par ailleurs des pistes pour construire des parades aux attaques désynchronisantes spécifiques à W-interp. Enfin, nous évoquons les liens entre le travail effectué sur l'interpolation et une récente technique de resynchronisation après attaque de torsion aléatoire.

C.1 Exemples de tatouages : étude subjective

Les figures suivantes montrent différentes images tatouées à DWR=25 dB (PSNR=40.5 dB). Nous avons choisi d'étudier un détail de l'image Lena. Il comporte essentiellement des zones planes et des contours, afin de mettre en évidence les points faibles des masques perceptuels. En effet, ce sont les zones où une modification de l'image est la plus perceptible. Dans les zones texturées, les masques possèdent d'excellentes propriétés perceptuelles. Ce DWR est légèrement au-dessus de la limite habituelle d'imperceptibilité, mais le détail est agrandi pour les besoins de l'étude.

Une étude subjective sommaire met en évidence les propriétés et les défauts des divers masques. Les différences entre images peuvent disparaître sur la version imprimée de ce rapport de thèse.

- DS sans masque crée un bruit, visible sur les zones planes de l'images (front et nez de Lena)
- pour le masque dans le domaine de la DCT, cette agitation est présente mais moins visible
- le masque NVF ne présente pas non plus d'artefact particulier (on voit une légère déformation de la mèche)
- les masques bilinéaire et scaling fournissent une bonne qualité perceptuelle, à part de légers artefacts au niveau de la mèche de cheveux
- le résultat d'une interpolation bilinéaire dégrade l'image (légère agitation sur les zones planes, accentuation des arêtes), mais ici DWR=18,5 dB
- lorsqu'on pondère directement l'erreur d'interpolation pour atteindre DWR=25 dB, ces artefacts sont peu visibles. Cependant, il est difficile de construire une technique de tatouage utilisant cette pondération directe
- lorsqu'on insère directement des erreurs d'interpolation, mais sur 25% des points uniquement, comme dans W-interp, on constate l'apparition d'artefacts : les zones de transition (mèche de cheveux) sont "tachetées". Pour les besoins de l'illustration, nous avons choisi une partie de l'image présentant des zones planes et des transitions abruptes. Ces artefacts n'apparaissent pas sur les zones texturées
- l'utilisation d'une interpolation par B-spline bicubique (algorithme de type W-spline) offre une meilleure qualité perceptuelle

La technique W-interp peut donc présenter des défauts perceptuels, exposés par une analyse subjective de l'image (mais non par les mesures perceptuelles objectives, cf. paragraphe 4.5). Nous nous sommes concentrés dans ce rapport sur une analyse théorique et objective du schéma de tatouage générique. Cependant, de nombreuses pistes, déjà évoquées, existent pour améliorer l'imperceptibilité : interpolation préservant les contours, autres techniques d'interpolation, autres grilles d'interpolation, limitation de la distorsion maximale...

original

DS simple

DCT

NVF

masque scaling

masque bilineaire DS

masque interp total

masque interp pondere

masque W−bilin

masque W−spline

C.2 Estimation itérative du seuil empirique

Lien entre seuil théorique et seuil empirique

Le calcul du seuil empirique est inspiré de la statistique de test $\mathbf{T} = \sum_{\mathcal{S}} r_k^2 \lessgtr \eta$. Cependant, $\eta = \frac{1}{L}\sum_{l=1}^{L} \rho_l^2$ ne converge pas vers η_{th} lorsque L augmente, ce qui souligne l'intérêt d'utiliser un seuil théorique. En effet,

$$\eta = \frac{1}{L}\sum_{l=1}^{L} \rho_l^2 = \frac{1}{2}\left(\frac{1}{L}\sum_{l=1|m_l=-1}^{L} \rho_l^2 + \frac{1}{L}\sum_{l=1|m_l=+1}^{L} \rho_l^2\right)$$
$$\longrightarrow \frac{1}{2}(\sigma_{\mathbf{n}}^2 \mathrm{E}[\chi_P^2] + (\sigma_{\mathbf{n}}^2 + \sigma_{\epsilon(\mathbf{x})}^2)\mathrm{E}[\chi_P^2]) = P(\sigma_{\mathbf{n}}^2 + \frac{1}{2}\sigma_{\epsilon(\mathbf{x})}^2)$$

ce qui ne correspond jamais à η_{th} (la différence s'atténue lorsque $\sigma_{\mathbf{n}}^2$ augmente).

Décodeur itératif proposé

Dans le scénario où l'attaque n'est pas connue (et sans se limiter au cas AWGN), on peut améliorer le seuil empirique en reformulant le décodeur de W-interp comme un détecteur de variance : le test s'effectue entre deux hypothèses

Hypothèse H_1 : présence d'un tatouage

$$\sigma_{\mathbf{r}|H_1}^2 = \sigma_{\epsilon(\mathbf{n})}^2$$

Hypothèse H_0 : absence de tatouage

$$\sigma_{\mathbf{r}|H_0}^2 = \sigma_{\mathbf{n}}^2 + \sigma_{\epsilon(\mathbf{x})}^2$$

Le seuil de décision optimal peut être calculé en connaissant $\sigma_{\mathbf{n}}^2$. Ici, on veut calculer une estimation $\hat{\epsilon}_{\mathbf{r}|H_0}$ et $\hat{\epsilon}_{\mathbf{r}|H_1}$ des variances de $\mathbf{r}$ sous les deux hypothèses, ce qui permettra d'estimer $\sigma_{\mathbf{n}}^2$. L'algorithme est donc le suivant (cf. *fig.* C.1) :

- itération 0 : $\eta^{(0)} = \frac{1}{L}\sum_{l=1}^{L} \rho_l^2$ (initialisation au seuil empirique)
- itération $t > 0$:
  - calculer l'estimation de $\mathbf{m}$ avec le seuil $\eta^{(t-1)}$:
    $\hat{m}_l^{(t)} = +1 \quad \text{si} \quad \rho_l^2 < \eta^{(t-1)}, \quad \hat{m}_l^{(t)} = -1 \quad \text{sinon}$
  - calculer l'estimation de $\sigma_{\mathbf{n}}^2$ correspondant à $\mathbf{m}$:
    $(\hat{\sigma}_{\mathbf{r}|H_0}^{(t)})^2 = \mathrm{Var}[r_k | b_k^{(t)} = 0]$, $(\hat{\sigma}_{\mathbf{r}|H_1}^{(t)})^2 = \mathrm{Var}[r_k | b_k^{(t)} = +1]$
  - calculer $\eta^{(t)}$, seuil de décision optimal associé aux variances $\hat{\epsilon}_{\mathbf{r}|H_0}^{(t)}$ et $\hat{\epsilon}_{\mathbf{r}|H_1}^{(t)}$

Le seuil empirique itératif est donc construit dans un souci d'optimalité du décodage.

Etude expérimentale

Les *fig.* C.2 à C.5 montrent que le seuil de détection itératif apporte une amélioration significative du TEB par rapport au seuil empirique, en particulier face à la compression JPEG et l'attaque AWGN (où le seuil itératif coïncide presque avec le seuil

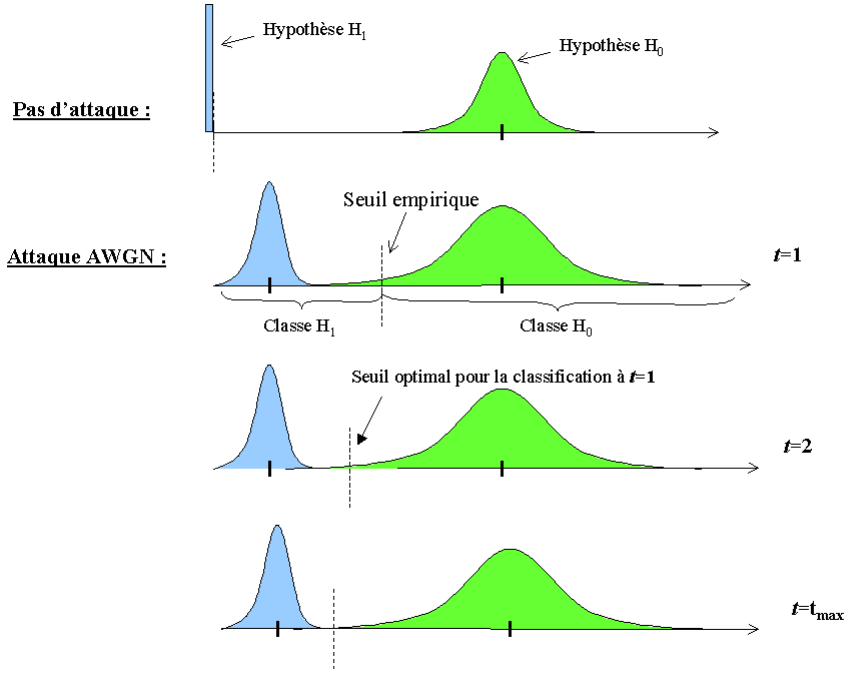

FIG. C.1 – Principe du décodage itératif proposé

optimal). L'amélioration est moins nette face à l'égalisation d'histogramme. La *fig.* C.6 montre que le décodeur itératif améliore considérablement la robustesse à la compression JPEG. Le décodeur itératif est même indispensable pour conserver l'invariance à une attaque de gain (cf. *fig.* C.7).

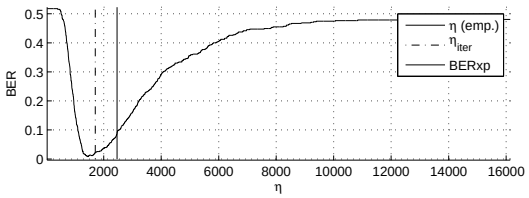

FIG. C.2 – Seuil itératif face au AWGN, DWR=28 dB, DNR=20 dB, Bateaux, $L = 512$

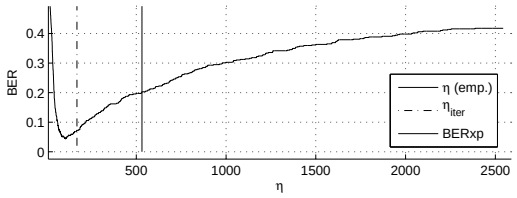

FIG. C.3 – Seuil itératif face au débruitage, DWR=28 dB, Bateaux, $L = 512$

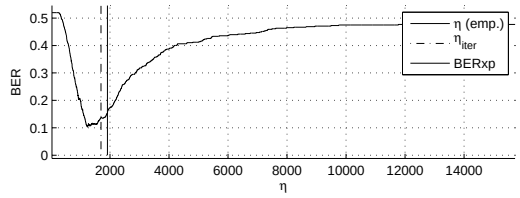

FIG. C.4 – Seuil itératif face à la compression JPEG, $Q = 85$, DWR=28 dB, Bateaux, $L = 512$

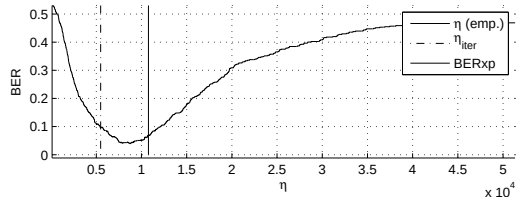

FIG. C.5 – Seuil itératif face à l'égalisation d'histogramme, $Q = 85$, DWR=28 dB, Bateaux, $L = 512$

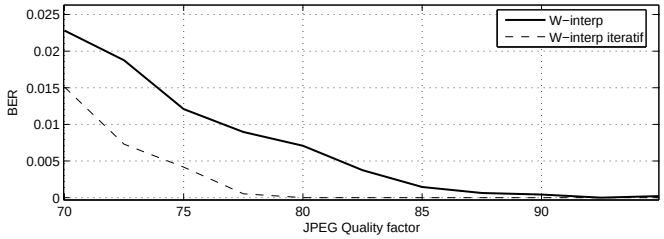

FIG. C.6 – Robustesse de W-bilin avec seuil itératif à la compression JPEG, $L = 64$, DWR=28 dB

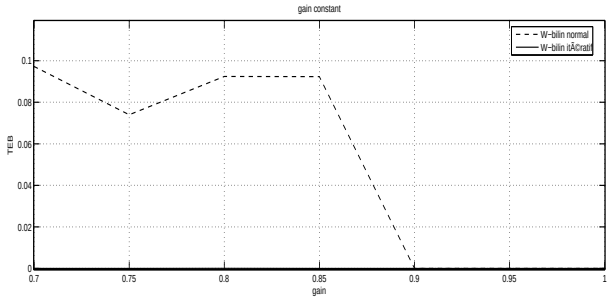

FIG. C.7 – Comparaison entre seuil itératif et seuil normal, attaque de gain constant, DWR=28 dB, $L = 1000$

C.3 Combinaison de W-interp et d'une technique d'optimisation

Dans sa version de base, W-interp n'est pas un algorithme d'insertion informée car l'insertion n'utilise pas la connaissance de l'algorithme de décodage. Cependant, W-interp peut être couplée avec une technique d'optimisation sur les paramètres du tatouage à l'insertion afin d'améliorer la robustesse et/ou l'imperceptibilité, un peu à la manière des techniques évoquées dans le paragraphe 1.2.8. Il s'agit alors d'**insertion informée**. On peut optimiser par exemple la robustesse à la compression JPEG, au débruitage, ou à une combinaison d'attaques déterministes. La technique peut également s'étendre à l'imperceptibilité, en optimisant la qualité SSIM par exemple. L'espace de recherche couvre les paramètres de W-interp, mais ne peut pas couvrir la clé $\mathbf{k}$ qui est secrète et connue du décodeur.

Il y a de nombreuses possibilités pour le choix de la technique d'optimisation.

Les algorithmes génétiques constituent cependant un bon candidat car l'espace de recherche est important, et donc difficile à parcourir de manière exhaustive. Il n'existe pas d'algorithme déterministe adapté et de complexité raisonnable. De plus, on cherche une bonne solution plutôt que la solution optimale mais difficile à obtenir. L'utilisation d'algorithmes d'optimisation et d'apprentissage pour sélectionner les meilleures marques respectant des critères donnés a déjà été développée dans le cadre du tatouage (cf. paragraphe 1.2.8). Huang et Wu [HW00], Shieh *et al.* [SHWP04] utilisent les algorithmes génétiques pour optimiser les lieux d'insertion du tatouage dans le domaine de la DCT. Kumsawat et Attakitmongcol [KA05] ont également proposé d'optimiser à l'aide d'algorithmes génétiques la qualité visuelle et la robustesse d'une méthode de tatouage dans le domaine de la DWT. La population optimisée est un ensemble de seuils de puissance d'insertion. Ils fournissent une étude de l'influence des divers paramètres : probabilités de croisement et de mutation, nombre de chromosomes, nombre de générations. Les chromosomes optimisés ici sont les paramètres du tatouage : puissance d'insertion, seuils d'insertion et de décodage, dans chaque sous-bande. L'approche proposée ici est donc plus proche de celle de Huang et Wu [HW00] et Shieh *et al.* [SHWP04].

Il est cependant difficile de calculer ou d'estimer les performances théoriques de la méthode proposée. De plus, le réglage des paramètres d'un algorithme génétique ne peut se faire qu'expérimentalement et nécessite une étude complète. Nous ne proposons donc pas ici de résultats expérimentaux.

C.4 Attaque intelligente sur la robustesse et modèle d'image

C.4.1 Attaques intelligentes sur la robustesse

Cas $\{\mathcal{S}_l\}_{l=1,\dots,L}$ connu, $K = \mathbf{g}$

Pour une image $\mathbf{x}$, pour un point (k_1, k_2) donné mais avec des coordonnées $(k_1 + \tau^u, k_2 + \tau^v)$ choisies aléatoirement, la différence entre deux interpolations bilinéaires suit une gaussienne $\mathcal{N}(0, \sigma^2_{\epsilon(\mathbf{s})})$. Cela représente l'erreur commise par le pirate lorsqu'il ignore $\mathbf{g}$. Elle est un peu inférieure à l'erreur d'interpolation $\sigma^2_{\epsilon(\mathbf{x})}$, et c'est la comparaison des deux qui fournit le niveau de sécurité. On montre alors que $\sigma^2_{\epsilon(\mathbf{s})} = \frac{7}{8}\sigma^2_{\epsilon(\mathbf{x})}$. En l'absence d'attaque, on a dans les deux hypothèses :
H_0, cas tatoué : au lieu de comparer la valeur interpolée avec le résultat de la même interpolation, on la compare avec le résultat d'une interpolation en des coordonnées différentes. On a donc un bruit : $r \sim \mathcal{N}(0, \sigma^2_{\epsilon(\mathbf{s})})$.
H_1, cas non tatoué : au décodage, on compare une interpolation avec des coordonnées quelconques avec la valeur d'origine. Donc $r \sim \mathcal{N}(0, \sigma^2_{\epsilon(\mathbf{x})})$. La sécurité augmente avec DWR (donc lorsque P diminue).

Attaque d'effacement utilisant l'interpolation

Comme W-interp est substitutive, elle est vulnérable à une substitution par une valeur quelconque, comme c'est le cas de la méthode LSB (cf. paragraphe 1.2.7). Deux problèmes se posent au pirate : 1. extraire les points à substituer ; 2. respecter la contrainte d'imperceptibilité (moins contraignante pour le pirate cependant). Le

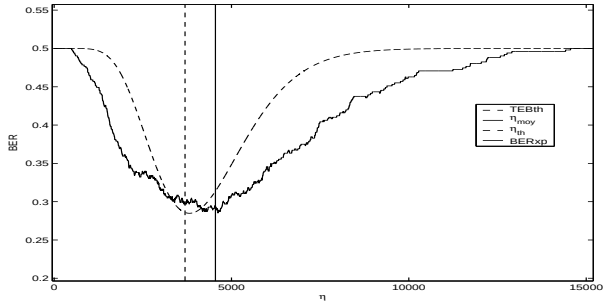

FIG. C.8 – Exemple de TEB lorsqu'on ignore la clé, W-interp, Babouin, DWR=26

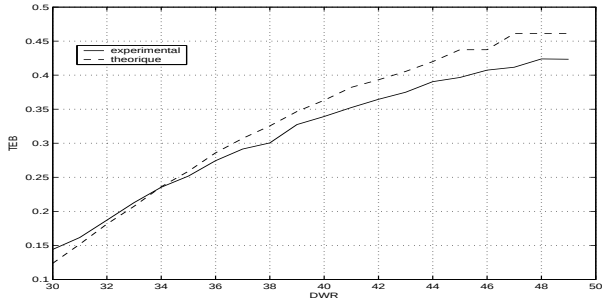

FIG. C.9 – TEB en fonction de DWR **g** inconnu, W-interp, Babouin, $L = 512$

point 1. est irréalisable, le pirate devant faire une substitution systématique, de puissance $\sigma^2_{\mathbf{n}} = \frac{\sigma^2_{\mathbf{x}}}{\text{DNR}}$. Si DWR=DNR, $\sigma^2_{\mathbf{n}} = \frac{N_S}{N}\sigma^2_{\epsilon(\mathbf{x})}$ (mais $\mathbf{n}$ peut être proportionnel à $\epsilon(\mathbf{x})$). W-interp est d'autant plus protégé de cette attaque si $\sigma^2_{\epsilon(\mathbf{x})}$ est grand et N_S petit. Des simulations montrent que W-interp n'est pas du tout affecté par une telle attaque (seulement 15% des points tatoués sont touchés, à DWR=28 dB pour Lena).

C.4.2 Utilisation d'un modèle d'image

Ce paragraphe a pour but de calculer théoriquement $\sigma^2_{\mathbf{s}}$, qui détermine le niveau de sécurité si $\{\mathcal{S}_l\}_{l=1,\ldots,L}$ est connu. Pour ce faire, il faut choisir un modèle d'image naturelle. Nous voulons modéliser un parcours causal de l'image sur les colonnes par une chaîne de Markov [SMCM05]. De même, on pourrait définir d'autres sens de parcours : causal sur les lignes, anticausal sur les lignes ou les colonnes. On dit que $\mathbf{x}$ est une chaîne de Markov si et seulement si

$$\forall \quad (k_1, k_2) \quad \text{p}[x_{k_1,k_2} = k | \mathbf{x} \setminus x_{k_1,k_2}] = \text{p}[x_{k_1,k_2} = k | x_{k_1-1,k_2}]$$

$\mathbf{x}$ est une chaîne de Markov-Gauss si et seulement si

$$\text{p}[x_{k_1,k_2} = k | x_{n_1-1,n_2}] = \frac{1}{\sigma_U\sqrt{2\pi}}\exp(-\frac{(k - x_{n_1-1,n_2})^2}{2\sigma_U^2})$$

Soit $n_{ij}(\mathbf{x})$ le nombre de transitions de la valeur i d'un site de l'image $\mathbf{x}$ vers la valeur j en un site voisin. On additionne les transitions sur les lignes et sur les colonnes, dans le sens causal ou anticausal, ce qui est cohérent puisque ces relations de voisinage sont utilisées dans W-interp. La matrice de co-occurrence, qui est un estimateur de la densité de probabilité conjointe, est $C(\mathbf{x}) = (n_{ij}(\mathbf{x})/N)$. Si les données suivent un modèle de Markov-Gauss, **c** sera donc une matrice creuse dont les éléments sont regroupés autour de la diagonale (l'étalement autour de la diagonale dépendant de

σ_U), cf. *fig.*C.10. Pour une image naturelle, les transitions semblent plutôt suivre une distribution gaussienne généralisée (cf. *fig.* C.11).

FIG. C.10 – Matrices de co-occurrences de 3 images naturelles

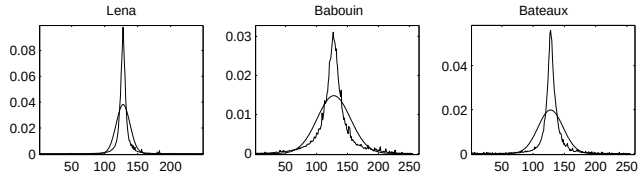

FIG. C.11 – Estimation de la distribution des transitions et gaussienne associée : $\sigma_U = 10.45, 26.9, 20$ respectivement

$$\epsilon(x_k) = (\sum_{j=1}^{N_v} g_j \underline{x}_j) - x_k = \sum_{j=1}^{N_v} g_j (\underline{x}_j - x_k)$$

Donc $\mathrm{E}[\epsilon(X)] = 0$ et

$$\begin{aligned}\mathrm{Var}[\epsilon(\mathbf{x})] &= \mathrm{E}[\left(\sum_{j=1}^{N_v} g_j(\underline{x}_j - x_k)\right)^2] \simeq \mathrm{E}[\sum_{j=1}^{N_v} g_j^2 (\underline{x}_j - x_k)^2] \\ &= \sum_{j=1}^{N_v} \mathrm{E}[g_j^2]\mathrm{E}[(\underline{x}_j - x_k)^2] = \mathrm{E}[\sum_{j=1}^{N_v} g_j^2]\sigma_U^2 \\ &= \Delta\sigma_U^2\end{aligned}$$

On a supposé dans ce calcul que les variables $(\underline{x}_j - x_k)$ étaient indépendantes, ce qui n'est pas le cas. On a donc négligé les termes croisés dans la somme. On peut également calculer, dans le cas particulier de W-bilin, $\sigma^2_{\epsilon(\mathbf{s})}$ en cas d'erreur sur la valeur des décalages, utilisée pour la sécurité : $\mathrm{E}[\mathbf{s}] = 0$ et

$$\begin{aligned}\mathrm{Var}[\mathbf{s}] &= 4\mathrm{E}[U^2](\mathrm{E}[((\frac{1}{2} \pm \tau^u)(\frac{1}{2} \pm \tau^v))^2] + \mathrm{E}[((\frac{1}{2} \pm \tau^{x'})(\frac{1}{2} \pm \tau^{y'}))^2] \\ &- 2\mathrm{E}[(\frac{1}{2} \pm \tau^u)(\frac{1}{2} \pm \tau^v)(\frac{1}{2} \pm \tau^{x'})(\frac{1}{2} \pm \tau^{y'})]) \\ &= 4\sigma_U^2(2\frac{1}{9} - 2\frac{1}{16}) = \frac{7}{18}\sigma_U^2\end{aligned}$$

Et au bilan :

$$\sigma^2_{\epsilon(\mathbf{x})} = \frac{7}{8}\sigma^2_{\epsilon(\mathbf{x})}$$

ce qui correspond à ce qui avait été observé expérimentalement. Plus généralement,

$$\sigma^2_{\epsilon(\mathbf{x})} = 2\sigma_U^2(\Delta - \frac{1}{4})$$

pour toute distribution centrée de $\mathcal{T}$. $\frac{1}{4}$ est une borne inférieure de Δ. Plus Δ s'éloigne de cette borne, plus le gain en sécurité est important. Le même calcul est possible pour W-spline, qui obtiendrait un meilleur niveau de sécurité.

C.5 Parades aux attaques désynchronisantes pour W-interp

Dans cette section, on identifie des scénarios où une variante de W-interp plus robuste aux attaques géométriques peut être construite. Les cas où une amélioration peut être apportée sont les suivants :

- l'attaque est connue au décodage et on veut l'inverser
- l'attaque est connue à l'insertion : on veut pré-déformer $\mathbf{w}$
- on cherche une invariance à une transformation donnée

Un quatrième objectif serait de construire une variante de W-interp permettant d'estimer l'attaque. La structure de W-interp rend cependant difficile la construction de $\mathbf{w}$ possédant des propriétés de corrélation adéquates.

Les outils à notre disposition sont les suivants :

- l'interprétation géométrique des décalages τ_u, τ_v à l'insertion ou au décodage
- la possibilité d'appliquer une pré-déformation sur la grille $\mathcal{G}$
- la possibilité d'agir sur la structure de $\mathcal{G}$, en éloignant les pixels ou en introduisant une grille suivant une forme géométrique donnée

Pré-distorsion de la grille d'interpolation : si la transformation géométrique $\mathcal{A}$ est connue à l'insertion, on peut utiliser $y_k = g(\underline{\mathcal{A}(x)}'_k)$, où $\underline{\mathcal{A}(\mathbf{x})}'$ sont des échantillons d'une version de $\mathbf{x}$ attaquée par $\mathcal{A}$. Ainsi, on aura au décodage : $g(\underline{z}_k) \simeq g(\underline{x}'_k)$. Les points de $\mathcal{G}$ doivent de plus être éloignés entre eux (ce qui nuit à l'imperceptibilité), afin de limiter l'impact de $\mathbf{w}$ sur $\underline{z}$.

Grille d'interpolation invariante : la nature géométrique de la grille $\mathcal{G}$ permet d'envisager d'utiliser par exemple, des cercles concentriques pour résister à une rotation ou un segment pour résister à une translation. Si le voisinage est invariant, on peut ainsi décoder en effectuant une recherche exhaustive. L'effet de la désynchronisation disparaît, mais pas celui du bruit d'interpolation.

Mise en forme invariante : une idée naïve est de construire des ensembles d'insertion $\mathcal{S}_l$ invariants par rapport à une attaque géométrique donnée : par exemple, des cercles concentriques ou des segments parallèles. La détection consiste alors à différencier les ensembles tatoués ou non. Cette technique nuit cependant à la sécurité et à l'imperceptibilité.

Correction de l'attaque ou canal d'attaque connu au décodage : dans certains scénarios on connaît la grille de déformation au décodage (exemple : impression puis passage par un scanner). Dans ce cas, nous proposons de décoder en interpolant sur la grille déformée elle-même. Cette méthode pourrait donner de bons résultats si on connaissait la valeur de l'image sur la grille déformée (continue). Par exemple, on pourrait annuler l'attaque grâce à des techniques d'interpolation inversibles. En réalité, on n'a cependant accès qu'à une version projetée sur $\mathbb{Z}^2$ de l'image déformée. On ne peut donc pas appliquer la méthode précédente. Il faut alors reconstruire les valeurs de l'image sur la grille déformée à partir des valeurs sur $\mathbb{Z}^2$. Ceci ne peut se faire que par interpolation. Au bilan, on a donc 2 interpolations successives, ce qui revient à effectuer $\mathcal{A}^{-1} \circ \mathcal{A}$. On se ramène donc à une méthode de resynchronisation classique.

Insertion dans des caractéristiques invariantes de l'image : une solution très différente est d'utiliser l'invariance des contours de l'image. Nous proposons donc d'extraire les contours de $\mathbf{x}$, puis de les tatouer par interpolation (par exemple : interpolation 1D en alignant les points successifs). Cette technique est cependant très éloignée de la version de base de W-interp.

C.6 Lien entre la sécurité de W-interp et les techniques de resynchronisation

Attaque de torsion aléatoire

L'attaque de torsion aléatoire (*random bending*), notamment utilisée dans le logiciel StirMark, consiste à appliquer des transformations géométriques locales non affines à $\mathbf{y}$. Son modèle est le suivant [BEH02] :

$$\begin{aligned} z_k &= \sum_j y_j g((k-j)T_a + \tau_j)T_a + n_k \\ &= y_k g(\tau) + \sum_{j \neq k} y_j g((k-j)T_a + \tau_j)T_a + n_k \end{aligned} \tag{C.1}$$

g étant une fonction d'interpolation, T_a une période d'échantillonnage, $\mathbf{n}$ un bruit additif gaussien, et τ les paramètres de torsion.

Resynchronisation par PLL

Afin de resynchroniser $\mathbf{z}$, on se place dans le cas où l'on connaît g, et on veut estimer τ. Dans [WBSH06], un algorithme de resynchronisation utilisant une boucle de phase (PLL) est proposé dans le cas 1D pour l'algorithme DM et une interpolation par sinC. L'intérêt de la PLL est de s'appliquer au cas où τ est constant, mais également au cas où τ varie doucement. Par exemple, [WBSH06] prend le modèle de la "marche aléatoire", qui correspond à l'attaque de torsion aléatoire : $\tau_k = \sum_{j=1}^{k} g_j$ où $g_j \sim \mathcal{N}(0, \sigma_{\mathbf{g}}^2)$. Le principe de cette PLL est le suivant : soit $\hat{\tau}_k$ l'estimation de τ_k à la réception, $\epsilon_k = \tau_k - \hat{\tau}_k$ l'erreur d'estimation et $\hat{\epsilon}_k$ l'estimation de cette erreur d'estimation.

$$\hat{\tau}(k+1) = \hat{\tau}_k + \nu \hat{\epsilon}_k$$

où ν est un paramètre de gain, compromis entre l'agilité de la boucle de poursuite et l'atténuation du bruit dans $\hat{\epsilon}_k$, et où

$$\hat{\epsilon}_k = \hat{y}_k y(k-1) - \hat{y}(k-1) y_k / 2\sigma_{\mathbf{y}}^2,$$

avec

$$\hat{y}_k = \sum^{j} z(j) h_A((k-j)T_a - \hat{\tau}(j))$$

On doit donc décider *a priori* de y, ce qui est possible pour DM en considérant les différents points de la grille.

Liens avec W-interp

La proximité entre l'équation (C.1) et le modèle de W-interp (4.4) soulève deux idées. Tout d'abord, on peut adapter la PLL de [WBSH06] à une estimation de la clé $\mathbf{g}$ de W-interp pour l'attaque KOA (et en supposant les messages connus). Pour l'attaque WOA, cet outil doit être combiné à une phase d'estimation de $\mathbf{b}$ et de $\mathcal{S}$. On se ramène donc à un algorithme EM similaire à celui proposé dans la partie 4.7, la PLL se substituant à la résolution du système linéaire. Dans le cas de W-interp et pour une attaque sur la sécurité, on ne connaît pas cependant le dictionnaire d'insertion, contrairement à la resynchronisation pour DM (on pourrait utiliser une estimation de $\mathbf{y}$ avec décalages

nuls). De plus, l'hypothèse de marche aléatoire est très restrictive sur le choix de la fonction $\mathbf{g}$ de W-interp. Du fait de ces restrictions et du passage en 2D pour l'application au tatouage d'images, nous n'avons pas implanté cette technique.

La seconde idée est d'utiliser l'algorithme EM de Popescu et Farid, que nous avions adapté à une attaque sur la sécurité de W-interp, à un problème de resynchronisation face à la torsion aléatoire. Dans [PF05], on détecte des traces de rééchantillonnage. Sans compression JPEG supplémentaire, on peut identifier quelles images ont été attaquées. Dans le cas d'une resynchronisation, la difficulté est plus grande : on veut identifier les paramètres eux-mêmes de l'attaque. L'algorithme ne fonctionnera que si τ est constant, contrairement à la PLL précédente. La phase d'estimation peut servir à détecter des régions épargnées, ou à détecter l'usage de différents paramètres par région. La technique ne se limite pas au DM.

Bibliographie

[BEH02] R. Bauml, J.J. Eggers, and J. Huber. A channel model for watermarks subject to desynchronization attacks. *Proc. of SPIE*, 4675, 2002.

[HW00] C.-H. Huang and J.-L. Wu. A watermark optimization technique based on genetic algorithms. *Proc. SPIE*, 3971 :516–523, 2000.

[KA05] P. Kumsawat and K. Attakitmongcol. A new approach for optimization in image watermarking using genetic algorithms. *IEEE Trans. on Signal Proc.*, 53(12), 2005.

[PF05] A.C. Popescu and H. Farid. Exposing digital forgeries by detecting traces of resampling. *IEEE Trans. on Signal Processing*, 53(2) :758 – 767, 2005.

[SHWP04] C.-S. Shieh, H.-C. Huang, F.-H. Wang, and J.-S. Pan. Genetic watermarking based on transform-domain techniques. *Pattern Recognition*, 37(3) :555–565, 2004.

[SMCM05] K. Sullivan, U. Madhow, S. Chandrasekaran, and B. S. Manjunath. Steganalysis of spread spectrum data hiding exploiting cover memory. *Proc. SPIE*, pages 38–46, 2005.

[WBSH06] K. Whelan, F. Balado, G. Silvestre, and N. Hurley. PLL-based synchronization of dither modulation data hiding. *Proc. of ICASSP*, 2006.

Publications associées à cette thèse

1. Vincent Martin, Marie Chabert, Bernard Lacaze. *Digital watermarking of natural images based on LPTV filters.* In : *IEEE Int. Conf. on Acoust., Speech and Sig. Proc.* ***(ICASSP'07)***, 2007.
2. Vincent Martin, Marie Chabert, Bernard Lacaze. *Substitutive watermarking algorithms based on interpolation.* In : *European Signal and Image Processing Conference* ***(EUSIPCO'06)***, Florence, Italy, September 4-8-2006. EURASIP.
3. Vincent Martin, Marie Chabert, Bernard Lacaze. *A novel watermarking scheme based on bilinear interpolation for digital images.* In : *IEEE Int. Conf. on Acoust., Speech and Sig. Proc.* ***(ICASSP'06)***, Toulouse, France, May 15-19-2006. IEEE.
4. Vincent Martin, Marie Chabert, Bernard Lacaze. *Single and multiple spread spectrum watermarking based on periodic clock changes.* In : *European Signal and Image Processing Conference* ***(EUSIPCO'05)***, Antalya, Turkey, September 4-8-2005. EURASIP.
5. Vincent Martin, Marie Chabert, Bernard Lacaze. *A Spread Spectrum Watermarking Scheme based on Periodic Clock Changes for Digital Images.* In : *7th Information Hiding Workshop* ***(IHW'05)***, Barcelona, Spain, June 6-8-2005. Universitat Oberta de Catalunya, pp. 67-81. LNCS.
6. Vincent Martin, Marie Chabert, Bernard Lacaze. *Un algorithme de tatouage d'images numériques reposant sur les changements d'horloge périodiques.* In : *GRETSI Symposium on Signal and Image Processing* ***(GRETSI'05)***, Louvain-la-Neuve, Belgique, September 6-9-2005. CNRS, SEE, pp. 1033-1036.
7. Vincent Martin, Marie Chabert, Bernard Lacaze. *Stratégies d'insertion informée pour un algorithme de tatouage utilisant l'interpolation bilinéaire.* In : *COmpression et REprésentation des Signaux Audiovisuels* ***(CORESA'06)***, Caen, France, November 9-10-2006.
8. Vincent Martin, Marie Chabert, Bernard Lacaze. *Introduction au Watermarking et à l'étalement de spectre par changement d'horloge périodique.* In : *Colloque de l'Ecole Doctorale d'Informatique et Télécommunications* ***(EDIT'04)***, Toulouse, France, 2004