Pierre Le Chapelain
DESS Cryptologie, Sécurité et Codage de l'information
Année 2002-2003

Rapport de stage

Analyse stéganographique d'images numériques. Comparaison de différentes méthodes.

Responsable de Stage : Patrick Bas
Tuteur : Roland Gillard

Date de soutenance : 23 juin 2003
Président du Jury : Franck Leprévost

Laboratoire des Images et des Signaux
961, rue de la Houille Blanche
BP 46, 38402 Saint Martin d'Hères

Table des matières

A Introduction

A.1 Intitulé du stage

La stéganographie consiste en la transmission d'un message secret dans un contenu quelconque sans que la présence du message soit décelable. La stéganalyse consiste à étudier les failles d'une méthode de stéganographie pour pouvoir attester que l'image contient ou non une information cachée.

L'objectif de ce stage est d'une part d'implémenter et de tester des méthodes de stéganographie couramment utilisées et d'autre part d'implémenter les algorithmes de stéganalyse connus dans la littérature et de concevoir de nouveaux outils permettant la stéganalyse de schémas existant. Ce stage est effectué au Laboratoire des Images et des Signaux de Grenoble (LIS).

A.2 Contexte du projet

A.2.1 Le LIS

Le LIS est concerné par les activités suivantes :
- Le traitement et l'interprétation des signaux et des images.
- La modélisation physique et l'expérimentation.
- Les applications à des situations réelles.

Son activité s'appuie sur de nombreuses coopérations dans le pôle grenoblois et à l'échelle nationale, européenne et internationale. Il participe au développement des connaissances par l'enseignement et la formation d'ingénieurs et de techniciens de haut niveau et par la diffusion de ses travaux de recherche dans la communauté internationale.

A.2.2 Le stage

Ce projet peut-être vu comme un travail préliminaire à la branche technique d'un projet plus vaste impliquant le LIS et trois partenaires : le CERDI (Centre d'Etudes et de Recherche en Droit de l'Immatériel attaché à la Faculté Jean Monnet de Paris XI), le projet TEMICS (Traitement, Modélisation d'Images & Communications basé à Rennes) et le LSS (Laboratoire des Signaux et Systèmes situé sur le site de Supelec à Gif-sur-Ivette) qui sont réunis à l'occasion d'une ACI visant à développer l'étude des techniques stéganographiques (aussi bien d'un point de vue théorique très lacunaire aujourd'hui, que du point de vue de ses applications) en synergie avec des juristes spécialistes des questions de propriété intellectuelle.

A.2.3 Calendrier

Après une phase de prise en main des outils du LIS disponibles, et un approfondissement préliminaire des notions théoriques, le projet se déroula en deux temps, chacun ayant pour objet l'étude d'un schéma différent de stéganographie. Leur étude est organisée suivant la même structure : un état de l'art en matière de stéganalyse du schéma est tout d'abord dressé, suivi par une phase d'implémentation et de tests qui s'achève par une analyse.

Tâches / Semaines	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Travail bibliographique	-	-	-													
Prise en main technique				-	-											
LSB : implémentation						-	-	-								
Tests & résultats									-	-						
QIM : implémentation											-	-				
Tests & résultats												-	-			
Rédaction											-	-	-	-	-	
Soutenance																-

B Contexte

B.1 Cryptographie et stéganographie

La sécurisation d'une communication entre deux personnes passe souvent par le cryptage du canal de transmission. Ce n'est cependant pas la seule méthode et celle-ci s'avère parfois inadaptée. Par exemple un message crypté intercepté entre un militaire et une ambassade d'un pays hostile prend du sens même sans connaître le contenu échangé et appelle la suspicion. Plus généralement, une communication sécurisée peut s'établir par cryptage de l'information, mais également par dissimulation du canal de transmission lui-même. L'étude de ce domaine inclut donc l'étude des techniques de dissimulation d'informations.

La stéganographie fait partie de ces techniques et consiste en la dissimulation d'un message à l'intérieur d'une information hôte, celle-ci devenant le canal de transmission, son médium. Ce message est inséré dans l'hôte en utilisant une clef de sorte qu'un tiers n'en ayant pas connaissance ne puisse pas détecter sa présence ni l'enlever. Tous les formats hôtes sont possibles et on pourra insérer tout type d'information dans un document texte, dans une image, dans une vidéo, etc.

La stéganographie n'est pas récente, on trouve trace de son utilisation chez les grecs et les perses en 500 avant J-C ou plus récemment en 1586 par exemple où elle fût utilisée dans la préparation de manœuvres conspiratrices visant le trône d'Angleterre, comme le relatent dans [26] les auteurs. Le champ de la stéganographie regroupe des pratiques qui peuvent être bien connues comme l'insertion de marques dans des lettres manuscrites à l'aide d'encre sympathique, ou l'utilisation de pochoirs ; on peut penser également à la lettre que George Sand écrivît à Alfred de Musset dont le vrai sens émergeait en ne lisant qu'une ligne de la lettre sur deux. Les techniques modernes de stéganographie concernent les formats numériques : quelques unes de celles-ci sont décrites dans [22] et [23].

Les méthodes de stéganographie utilisées seront fonctions du contexte de la communication et du type d'hôte. Bien entendu la stéganographie n'a de sens qu'en présence d'un ennemi, qui pourra selon les cas être actif ou passif s'il peut respectivement interférer dans le canal de transmission et modifier les informations, ou écouter le trafic y prenant place. La nature de cet ennemi influence donc également les techniques de stéganographie utilisées : les attaques du schéma de tatouage varient beaucoup entre ces deux cas et les notions de robustesse du schéma aux attaques et de capacité du canal de transmission permettent aujourd'hui de les classer.

B.2 Tatouage et stéganographie

Les techniques de tatouage ont aujourd'hui le vent en poupe, en raison de l'inquiétude croissante des producteurs de contenu numérique qui doivent, pour tirer bénéfice des nouveaux marchés engendrés par ces technologies, mettre en place des mécanismes de protection des œuvres distribuées. Le format numérique permet en effet des copies de masse à peu de frais et engendre autant de manque à gagner. Ainsi en est-il donc des distributeurs de musique, de films, de livres et de logiciels qui cherchent à inclure dans les fichiers vendus des marques numériques indécelables appelées tatouages ou filigranes (Watermark en anglais). Ces filigranes ont pour vocation la protection des droits d'auteur et sont classés en deux ensembles en fonction de leur contenu. Que celui-ci soit identique sur toutes les copies du fichier et l'on parlera de Watermarking. L'objectif - succédané d'un copyright - est alors de permettre d'identifier l'appartenance de l'œuvre au possesseur de la clef : en extrayant la marque de l'image à l'aide de sa clef secrète, le propriétaire fait la démonstration de ses droits sur le contenu. Que le filigrane soit différent pour chacune des copies et l'on parlera de Fingerprinting. L'objectif de ce marquage - équivalent d'un numéro de série - est de permettre au distributeur de l'œuvre d'identifier le contrevenant au contrat de licence.

Chacun de ces domaines de tatouage peut utiliser des méthodes différentes qui dépendent des fonctionnalités attendues et des attaques auxquelles ils seront confrontés et auxquelles ils résistent.

Les techniques de tatouage du watermarking et de la stéganographie peuvent sembler proches et dérivent en effet souvent des mêmes idées. Toutefois les utilisations foncièrement différentes font que les implémentations et adaptations des schémas et les questions soulevées par ces deux domaines sont radicalement différentes. Ainsi le watermarking est-il concerné principalement par la protection des droits d'auteur. Un enjeu majeur aujourd'hui est de pouvoir faire la preuve de l'efficacité d'un schéma de protection d'une œuvre numérique sur des périodes longues à l'échelle des progrès informatiques : les œuvres devraient en effet pouvoir être protégées pendant plusieurs décennies si besoin est. Cette question rejoint alors des questions de cryptographie plus générales où la démonstration scientifique rigoureuse de la sécurité d'un modèle indépendamment de la puissance des moyens de calcul devient essentielle.

La stéganographie est moins concernée par ces questions de durée. Elle est globalement soutenue par le besoin qu'ont certaines personnes de communiquer de façon sécurisée, communication par définition limitée dans le temps. Il peut s'agir de communication à des fins criminelles ou en environnement hostile (stéganographie sur des images par le biais d'internet, utilisation de mail anonyme (remailer)) mais également commerciale comme le contrôle automatique de publicité à la radio ou l'indexation de vidéos ou d'images médicales en insérant les commentaires directement dans le support numérique. Dans le domaine militaire, la technique stéganographique d'étalement de spectre

est par exemple utilisée pour les communications par radio. Ce type de communication peut suivre l'évolution des techniques d'analyse et s'adapter au gré des progrès de la stéganographie.

B.3 Attaques et stéganalyse

Le watermarking peut-être vu comme un tatouage en présence d'un ennemi actif. L'objectif de ce dernier est de perturber le signal hôte de manière à rendre toute détection du filigrane impossible. Ceci peut être possible sans connaître la présence de la marque. L'attaque consiste juste à la rendre inutilisable dans le cas où elle existe. Pour la stéganographie, une attaque peut-être jugée effective simplement lorsqu'elle détecte la présence du filigrane dans le signal, comme le notent les auteurs dans [24]. De telles analyses sont désignées par le terme spécifique de stéganalyse ou analyse stéganographique.

La stéganographie bénéficie aujourd'hui du même traitement scientifique que la cryptographie pour l'étude de sa robustesse. Les schémas de tatouage sont ainsi analysés en considérant que l'attaquant connaît l'algorithme de tatouage utilisé. La sécurité des schémas tiendra donc, dans le respect du principe de Kerckhoff, à la possession par l'émetteur d'une clef privée. Les stéganalyses tenteront de déceler les perturbations du signal hôte engendrées par la présence du filigrane.

C Stéganographie et stéganalyse sur des images

C.1 Domaines d'insertion

En toute généralité, les méthodes ou schémas de tatouage numérique sur des images sont communément classés par domaine d'insertion. Le domaine d'insertion est une représentation de l'image à tatouer que l'on modifie afin qu'elle porte la marque désirée. Il y a donc autant de domaines d'insertion que de systèmes de représentation des images. On a l'habitude d'en distinguer essentiellement trois (cf. [4]).

- Le domaine spatial qui correspond à l'espace "naturel" de représentation des images (formats pgm, bmp, etc.)
- Le domaine fréquentiel est le domaine de représentation des images après transformée TFD ou TCD (format JPEG par exemple) de l'image spatiale.
- Le domaine multirésolution, accessible après transformée en ondelettes de l'image (format JPEG2000).

C.2 Schémas

Chacun de ces domaines peut être utilisé comme espace d'insertion en leur appliquant un schéma d'une des deux grandes classes de tatouage : additifs ou substitutifs. Chacun des différents schémas et chacun des domaines aura ses points forts et ses points faibles variant en fonction de leur robustesse et des objectifs du tatouage (watermarking, fingerprinting, stéganographie).

Le tatouage additif est une méthode qui ajoute la marque à une composante de l'image. La technique du tatouage par étalement de spectre dans le domaine spatial en est par exemple un représentant et est décrite dans [4]. La détection s'effectue grâce au calcul d'une corrélation entre le signal aléatoire de modulation du message inséré et l'image tatouée.

Les tatouages substitutifs remplacent une composante de l'image à marquer par le filigrane ([4]).

C.2.1 QIM et DC-QIM

Les méthodes de quantification vectorielle spatiale proposées par Chen et Wornell ([9] et [10]) appartiennent à la classe des schémas substitutifs et consistent à remplacer des blocs de l'image par des blocs d'un dictionnaire prédéfini, choisi selon son maximum de vraisemblance avec le bloc qu'il remplace.

Soit $s(x, m)$ la fonction d'insertion du message, où x est le signal hôte et m le message à insérer. La distorsion induite par l'insertion doit être faible et on doit donc avoir l'égalité :

$$s(x, m) \approx x \ \forall m$$

La méthode QIM utilise les fonctions de quantification. Le principe d'insertion par quantification consiste d'abord à choisir deux fonctions de quantification (QF) non recouvrantes $\times$ et $\circ$. On quantifie alors le signal source x en fonction du message à insérer : si le message à insérer est un 0 on utilisera la fonction $\times$, sinon on utilisera la fonction $\circ$.

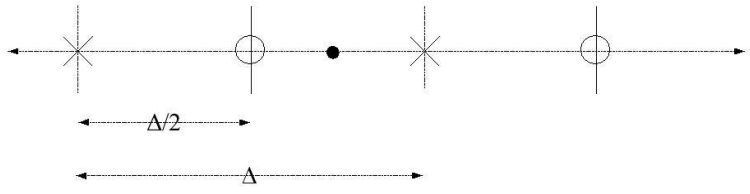

FIG. 1 – QIM : Le pixel • est quantifié en $\times$ ou ◯ selon la valeur du bit à insérer.

Δ est le pas de quantification. L'insertion d'un bit peut se décrire ainsi :

$$\begin{cases} si\ b = 0 & s(x,b) = \times(x) \\ si\ b = 1 & s(x,b) = \circ(x) \end{cases}$$

ou encore

$$\begin{cases} si\ b = 0 & s(x,b) = QF_\Delta(x) \\ si\ b = 1 & s(x,b) = QF_\Delta(x - \frac{\Delta}{2}) + \frac{\Delta}{2} \end{cases}$$

La détection est réalisée en évaluant l'inégalité de distance :

$$|s(x,m) - QF_\Delta(s(x,m))| < |s(x,m) - QF_\Delta(s(x,m) - \frac{\Delta}{2}) - \frac{\Delta}{2}|$$

Si celle-ci est vérifiée, alors le bit décodé vaut 0, sinon il vaut 1.

Un forme dérivée de quantification vectorielle utilise deux vecteurs de dilution d_0 et d_1 rendant le marquage plus robuste. La fonction d'insertion s'écrit alors :

$$s(x,m) = QF_\Delta(x + d(m)) - d(m)$$

Le vecteur de dilution d utilisé dépend du bit à insérer : d_0 pour un 0 à insérer et d_1 pour un 1. Le décodage procède mêmement, en évaluant la distance

$$|s(x,m) - QF_\Delta(s(x,m) - d_0(m)) - d_0(m)| < |s(x,m) - QF_\Delta(s(x,m) - d_1(m)) - d_1(m)|$$

Si celle-ci est vérifiée, alors le bit décodé vaut 0, sinon il vaut 1.

C.2.2 LSB

Le schéma LSB consiste à remplacer le bits de poids faible des pixels de l'image originale par le message à insérer. La perturbation visuelle engendrée par cette modification n'est en effet pas perceptible à l'oeil.

C.3 Etat de l'art

Les schémas de tatouage de type LSB furent très étudiés ces dernières années. Ce vocable est à prendre ici au sens large : il regroupe aussi bien les schémas de type LSB dans le domaine spatial (image en nuances de gris ou couleur en 24 bits) que les schémas de substitution à base de palette (EzStego sur les images GIF) ou dans le domaine fréquentiel (J-Steg ou Outguess qui manipulent les coefficients DCT des images JPEG). Pfitzman et Westfeld ([30] avaient ainsi proposé une méthode très efficace appellée attaque du χ^2, basée sur la détection des changements statistiques dans des paires de valeurs (valeurs de pixels, coefficients DCT ou indices de palette) engendrés par la présence d'un message.

Cette méthode ne donnait en revanche pas de bons résultats sur des messages insérés de façon aléatoire dans le plan de bits. Provos ([27]) proposa une attaque plus efficace, formalisée par Fridrich dans [17]. Ce même article contient la description d'une autre attaque basée sur des paires de valeurs et des calculs statistiques amenant à des seuils de détection très bas. Ces bons résultats sur tous les formats d'images sont cependant supplantés pour les schémas LSB sur les images couleurs 8-24 bits et en nuances de gris par une attaque antérieure proposée par Fridrich dérivé d'un schéma d'insertion proposé dans [12] : on trouvera la description d'une attaque sur les images couleurs dans [13] et la description de la Stéganalyse RS dans [16], [14]. Cette attaque est étudiée à la section E.5.

Cette avancée de la stéganalyse a engendré de nouveaux schémas d'insertion illustrant une fois de plus le jeu du chat et de la souris auxquels se livrent ces deux activités complémentaires que sont la stéganalyse et la stéganographie. Ces schémas concernent essentiellement le format JPEG, très étudié aujourd'hui car très répandu et utilisé avec des images naturelles, possédant une capacité supérieure aux images artificielles (pour une étude des formats adaptés on pourra consulter Aura [1], la notion de capacité stéganographique est abordée dans [21] et [8]. L'algorithme OutGuess a ainsi été proposé par Provos [28] pour contrer l'attaque du χ^2 et procède en deux temps pour insérer le message : insertion selon une ordre aléatoire, et correction afin de rendre l'histogramme de l'image marquée identique à l'image source. L'algorithme F5 pour les images JPEG a été présenté par Westfeld [29] en 2001 comme challenge à l'occasion du quatrième Information Hiding Workshop à Pittsburg. Il est également insensible au χ^2 ; il ne procède plus par permutation, mais par décrément d'un sur le LSB.

Ces deux schémas ont été tout dernièrement mis en échec par Fridrich ([18], [17], [21]), en tant que cas particuliers d'une attaque générale sur les systèmes stéganographiques appliqués aux images JPEG. Ces attaques reposent sur la modification d'une certaine donnée statistique macroscopique de l'image dans le processus d'insertion. Les détails de ces attaques sont décrits dans le cas de Outguess [18] et de F5 [20].

Parallèlement, Farid a proposé [11] une méthode de détection à l'aveugle basée sur des ensembles d'apprentissage et des décompositions des images en ondelettes et capable en théorie d'attaquer n'importe quel schéma d'insertion. L'analyseur proposé n'est en revanche pas capable aujourd'hui d'estimer la taille du message inséré, et est moins efficace que les analyses stéganographiques spécifiques aux schémas d'insertion.

Chandramouli et Memon ont proposé une stéganographie adaptative [7] s'opposant à l'attaque RS et permettant d'augmenter sensiblement la capacité d'insertion.

Avcibas et Memon présentent des stéganalyses sur les schémas LSB basées sur des mesures de qualité de l'image et de l'analyse de la variance multivariée en exploitant l'idée que la distance de l'image marquée à l'image bruitée est plus importante que la distance de l'image source à la même image bruitée ([2]). Une extension de ces analyses est donnée

dans [3] qui propose d'étudier les variations de certaines corrélations entre variables statistiques existant entre les différents plans de bits. Cette méthode permet d'attaquer des schémas d'insertion utilisant d'autre plans de bit que le dernier. Les auteurs rapportent ainsi de bons résultats sur Digimarc.

A coté des schémas de type LSB qui semblent bien traîtés, on citera des schémas émergeant ou peu étudiés : Chen et Wornell ont ainsi proposé en 1998 un schéma d'insertion substitutif QIM donnant de bons résultats mais n'ayant pas fait l'objet d'analyse suffisante ([10] et [9]). La section F présente les premiers résultats de notre analyse de ce schéma. Fridrich propose également dans un de ses derniers articles ([15]) un schéma basé sur de l'insertion de bruit qui généralise ce type de schémas additifs (cf. l'algorithme Hide par exemple) et Newman [25] propose un nouveau schéma d'insertion contrant l'attaque de Fridrich [21]. Des travaux plus théoriques ont également été menés par Cachin [5] et Chandramouli qui tente dans [6] de donner un cadre mathématique rigoureux à la stéganalyse. C'est également l'objectif poursuivit par J. J. Harmsen dans [19].

D Cadre d'étude

D.1 Périmètre

D.2 Organisation des tests

Les résultats présentés dans ce document se basent sur deux jeux d'images en nuances de gris sur 8 bits. Le premier jeu est constitué d'images classiques en traitement des images : Lena, Baboon, Water, TexMos et Peppers ont été utilisées dans leur version en taille 512x512 pixels convertis de leur format couleur d'origine à l'aide du logiciel convert sur une station Linux. Ces images ont été récupérés du site internet de Fabien Petitcolas ($http://www.cl.cam.ac.uk/\ fapp2/watermarking/image_database/index.html$). Le second jeu de tests est constitué d'images de taille 2960x1920 prises avec un appareil photo numérique et converties en nuances de gris de la même façon. Sappey, Latronche, Belledone et Futaie sont quatre photographies de paysage prises sans intention particulière. Ciel, Mur, Pelouse et SolPinede ont été prises afin de mettre en évidence l'influence pressentie de la texture sur les résultats des expériences. Ainsi Ciel et Mur sont-elles particulièrement peu texturées ; Pelouse et solPinede le sont fortement.

Les différences de taille entre les images permet d'obtenir des indications sur la précision des méthodes statistiques employées, les grandes images possédant naturellement plus de réalisations que les petites. Pour chacune de ces images, la production des données statistiques à l'origine de nos résultats provient de moyennes sur un ensemble d'expériences aléatoires répétées de dix à cinquante fois en utilisant le générateur de nombre pseudo-aléatoire de la librairie C standard, nourri avec autant de germes ou valeurs d'ini-

tialisation différents. Ceci nous a permis d'obtenir la régularité statistique des mesures indispensable à l'étude des phénomènes observés.

D.3 Remarques sur la programmation

L'ensemble des tests repose sur des calculs effectués à l'aide de primitives développées en C++. L'aspect prospectif du projet n'a pas permis d'exploiter pleinement les possibilités de la programmation orientée objet et l'essentiel du code produit s'appuie sur une bibliothèque de traitement des images interne au LIS, avec un fonctionnement de type C standard. Nous avons été de plus bien souvent contraints d'adapter les procédures pour obtenir un gain d'efficacité. La bibliothèque n'est en effet pas performante pour les calculs intensifs sur des images requis par nos tests. Il serait intéressant à l'avenir et dans la perspective des futurs projets du LIS de rationaliser le code produit et notamment de mettre en place des outils de tests efficaces en terme de temps machine.

Le code source du projet est distribué sur autant de librairies qu'il y eu de méthodes de stéganalyse testées adjointe d'une librairie d'outils regroupant des fonctions utiles à toutes et le code relatif au calcul du χ^2 (E.2.1). Ce dernier nécessite du calcul intégral approché et l'utilisation de la fonction gamma d'Euler. Leur implémentation fut intégralement tirée de l'ouvrage de référence *Numerical Recipes in C, The Art of Scientific Computing Second Edition* (chez CAMBRIDGE UNIVERSITY PRESS de William H. Press, Saul A. Teukolsky, William T. Vetterling et Brian P. Flannery).

E Stéganalyse : étude de l'existant

Cette section étudie en détail trois analyses stéganographique majeures issues des dernières avancées dans ce domaine. Ces trois méthodes sont analysées après deux parties introductives sur l'attaque visuelle et l'attaque du χ^2.

E.1 Introduction : attaques visuelles

L'insertion d'un message dans le dernier plan de bit peut se faire de façon aléatoire sur l'ensemble des pixels ou de façon séquentielle à partir du début de l'image. L'attaque visuelle décrite ici n'est pas efficace contre les méthodes d'insertion courantes, utilisant essentiellement une insertion aléatoire, ni sur les images très texturées. Toutefois cette attaque est une bonne introduction à la stéganalyse du schéma LSB en la rendant plus tangible et intuitive. Le principe de cette attaque est basé sur le fait que sur une image peu texturée, le plan LSB est corrélé avec l'image d'origine. L'insertion séquentielle du message perturbe le plan LSB en proportion de la taille du message. On observe donc, sur une coupe de ce plan une zone de bruit correspondant au message et une zone présentant des éléments de régularité non aléatoire (corrélés avec l'image d'origine). La figure 3 représente la coupe LSB de l'image suivante, dégradé vertical de 1280x960 pixels, avant et après insertion d'un message dans le plan LSB.

FIG. 2 – Dégradé : image originale

La coupe LSB de l'image initiale montre une régularité qui correspond à la régularité de l'image initiale : la corrélation entre le plan LSB et l'image est très forte. On remarque ainsi que la moitié supérieure de l'image tatouée est très bruitée et laisse apparaître de façon claire la présence d'un message dans l'image de taille 50%.

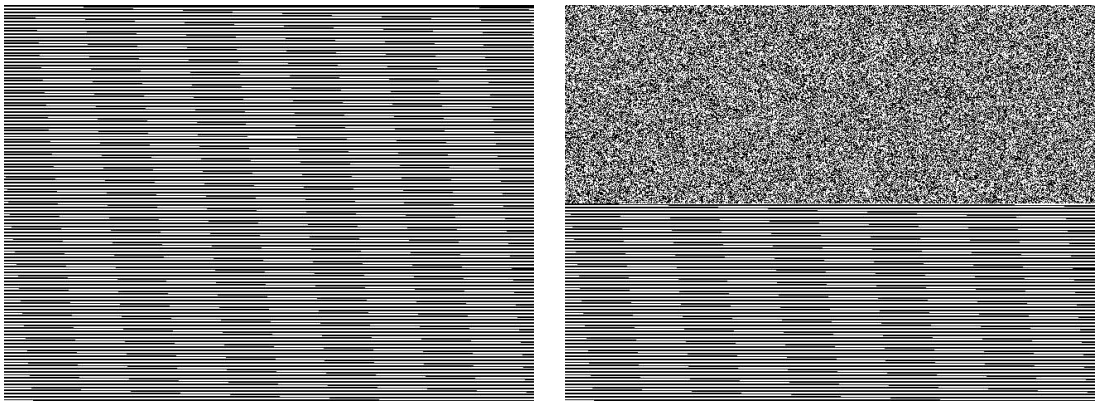

FIG. 3 – Dégradé : dernier plan de bits avant et après insertion

L'image Sappey de la figure 4, a été choisie car c'est une image naturelle qui possède des zones homogènes assez grandes et en assez grand nombre. En comparant les images de la figure 5, on peut encore identifier la présence d'un message (insertion d'un message de longueur 75% du plan de bit). Toutefois, on peut mesurer à l'aide de cette image la limite de l'attaque présentée : la décision de présence d'un message ne dépend plus ici que de peu de zones régulières. Le même test réalisé sur Lena par exemple ne révéla aucun artéfact identifiable à l'oeil.

FIG. 4 – Sappey : image originale

Bien que sans intêret pratique car présentant de trop mauvais résultats sur la majorité des images, cette attaque illustre une idée fondatrice des attaques étudiées dans la suite de ce document. Toutes ont pour objet la mesure d'une déviation d'une certaine quantité

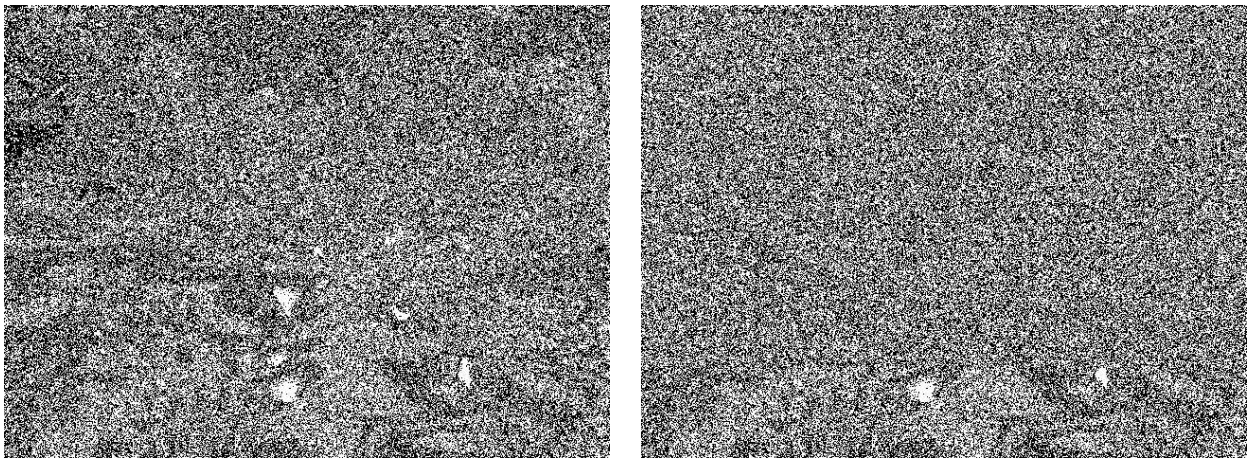

FIG. 5 – Sappey : dernier plan de bits avant insertion et après

statistique de l'image engendrée par la présence du message, qui s'apparente, comme on peut le voir avec [19] à l'ajout d'un bruit sur le plan LSB. Cette déviation porte sur des cardinaux d'ensembles de pixels dans les méthodes qui vont suivre.

E.2 Analyse statistique à base de χ^2

La stéganalyse basée sur l'analyse statistique des paires de valeurs de l'image est apparue avec les travaux de Westfeld et Pfitzmann ([30]) sur les tatouages LSB des images GIF. Nous l'avons adapté ici au cas de l'analyse du schéma LSB sur les images en nuances de gris sur 8 bits.

E.2.1 Rappels théoriques

Le test du χ^2 de Pearson fait partie de la famille des tests d'hypothèses. Si un répartition statistique est approchée par une courbe théorique $f(x)$, des écarts avec les mesures relevées sont inévitables, et ce même si la courbe théorique est bien choisie. Le test d'hypothèse consiste à évaluer si les écarts - étant donné le faible nombre d'observations - sont dus au hasard, ou au mauvais choix de la courbe de lissage f.

On cherche donc à vérifier l'hypothèse H selon laquelle une variable aléatoire X suit une certaine loi de répartition $F(x)$. On considère donc la grandeur U égale à la différence entre la répartition théorique et statistique, qui est une variable aléatoire dont la loi de répartition dépend de celle de X et du nombre d'expériences. La mesure U que nous avons choisie pour évaluer cet écart à pris une valeur u ; on cherche a savoir si cette valeur est trop importante pour s'expliquer par le hasard, ou si les répartitions théorique et

statistique sont assez proches. On suppose donc que l'hypothèse H est vraie et on veut calculer $P(U \geq u)$. En cas de probabilité faible on rejette l'hypothèse, sinon les données relevées ne sont pas en contradiction avec l'hypothèse.

On suppose que l'on ait réalisé n expériences dans lesquelles X a pris des valeurs dans k intervalles I_i. Si on connait la loi de répartition théorique, on connait la probabilité théorique p_i, $i \in [1..k]$ que la variable aléatoire X prenne une valeur située dans l'intervalle I_i. On choisit la mesure de non conformité U comme suit :

$$U = \chi^2 = \sum_{k\ catégories} \frac{(fréquence\ observée - fréquence\ théorique)^2}{fréquence\ théorique}$$

Si n_i $i \in [1..k]$ est le nombre de valeurs que X a prise dans l'intervalle I_i durant les n expériences, alors

$$\chi^2 = \sum_{j=1}^{k} \frac{(n_j - np_j)^2}{np_j}$$

On appelle loi du χ^2 à r degré de liberté la répartition des carrés de r variables aléatoires indépendantes, dont chacune est répartie suivant un loi normale d'espérance mathématique nulle et de variance unité. La densité de cette répartition est

$$k_r(u) = \begin{cases} \frac{1}{2^{\frac{r}{2}}\Gamma(\frac{r}{2})} u^{\frac{r}{2}-1} e^{-\frac{u}{2}} & pour\ u > 0 \\ 0 & pour\ u < 0 \end{cases}$$

avec

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$$

la fonction Γ d'Euler.

Si le plus petit des nombres n_j est supérieur ou égale à 5, la quantité χ^2 a pour distribution approchée la loi continue du χ^2 à $(k-1)$ degrés de liberté. On peut montrer que le nombre de degrés de liberté s'obtient en soustrayant le nombre de contraintes indépendantes sur les effectifs et le nombre de paramètres indépendants estimés à partir des données. On peut alors calculer la probabilité p pour que la variable répartie suivant la loi du χ^2 à r degré de liberté dépasse la valeur du χ^2 obtenue. La région d'acceptation du test est l'intervalle $[0, \chi_r^2]$ tel que la probabilité d'une variable du χ^2 à r degrés de liberté prenne une valeur dans cet intervalle soit égale à $1 - \alpha$ (α étant l'erreur de première espèce relative au test). Si la valeur de l'indicateur est supérieure à χ_r^2 , alors on décide l'hypothèse de non conformité, sinon on accepte l'hypothèse.

E.2.2 Principe de la stéganalyse

Le principe de la stéganalyse par χ^2 proposé par Westfeld et Pfitzmann se décrit comme suit. Les nuances de gris d'une image prennent leur valeurs dans l'ensemble

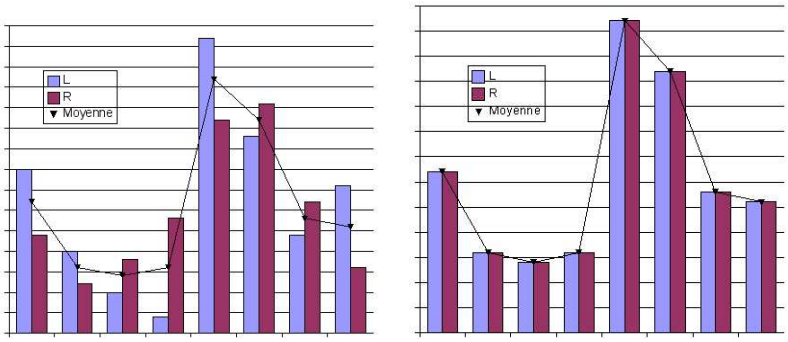

FIG. 6 – Partie d'un histogramme d'une image avant et après insertion LSB. Les fréquences d'apparition des nuances d'une paire de valeurs tendent à l'égalité sous l'action d'insertion LSB.

$P = \{0, 1, 2, ..., 255\}$. L'insertion d'un message dans le plan LSB modifie les fréquences de nuances adjacentes (voir les histogrammes en figure 6). Si n_i est la fréquence d'apparition de la nuance $i \in P$ dans l'image initiale et n_i^* la fréquence d'apparition de la nuance i dans l'image marquée, alors

$$|n_{2i} - n_{2i+1}| \geq |n_{2i}^* - n_{2i+1}^*|$$

L'insertion réduit l'écart de fréquence entre des nuances adjacentes au sens LSB, i.e. qui sont échangées l'une l'autre par le procédé d'insertion : $0 \leftrightarrow 1$, $2 \leftrightarrow 3$, . . . , $254 \leftrightarrow 255$. Les catégories de notre test sont donc les différentes paires de valeurs précédentes. Si leur occurence n'est pas assez forte (leur fréquence doit être supérieure à 4, cf. E.2.1), on devra regrouper plusieurs paires de valeurs ensembles. La fréquence théorique attendue dans la catégorie i après l'insertion LSB d'un message est

$$n_i^* = \frac{n_{2i} + n_{2i+1}}{2}$$

La fréquence mesurée sur l'image analysée est $y_i = n_{2i}$. Alors la valeur du χ^2 mesurant la différence des distributions est :

$$\chi^2_{\nu-1} = \sum_{i=1}^{\nu} \frac{(y_i - y_i^*)^2}{y_i^*}$$

où ν est le nombre de degrés de libertés qui correspondent ici au nombre de catégories de paires de valeurs utilisées. La probabilité que ces deux distributions soient égales est donné par le calcul suivant :

$$p = 1 - \frac{1}{2^{\frac{\nu-1}{2}}\Gamma(\frac{\nu-1}{2})} \int_0^{\chi^2_{\nu-1}} e^{-\frac{x}{2}} x^{\frac{\nu-1}{2}-1} dx$$

E.2.3 Résultats

Les résultats de cette stéganalyse sont décevants à deux titres. D'abord cette méthode ne donne aucune estimation de la taille du message. Ensuite les tests réalisés sur nos images n'ont pas donné de résultats probants. Ce test n'est efficace que pour des taux d'insertion élevés et même à ces taux l'erreur de détection est importante : le test du χ^2 sur Pelouse avec une taille de message de 100% n'a donné qu'une probabilité de présence de message de 4%.

Nous remarquerons en outre que cette analyse par χ^2 nécessite en toute rigueur que les catégories ou paires de valeurs considérées soient statistiquement indépendantes. Cette hypothèse est hautement improbable pour des images naturelles : l'histogramme de ces images n'est pas aléatoire, et une corrélation existe entre un niveau de gris et un autre.

E.2.4 Développements récents

On notera que dans [27], Provos a adapté la méthode précédente aux images JPEG, en effectuant le test du χ^2 sur les coefficients DCT de l'image. Constatant que l'insertion d'un message aléatoire produit des fluctuations du χ^2 au début, il proposa d'adapter la méthode en réalisant ce test sur une fenêtre de taille plus petite que l'image. Fridrich dans [17] a explicité cette technique ; toutefois les résultats obtenus ne permettent ni d'estimer la taille du message ni ne donnent de résultats satisfaisants sur tous les types d'images. D'autre part cette méthode de détection est très facile à contourner : Provos en donne le moyen lui-même dans son article avec l'algorithme OutGuess.

E.3 LIS

E.3.1 Principe

La méthode de stéganalyse du LIS est basée sur des égalités statistiques de cardinalités d'ensembles de paires de pixels qui sont modifiées sous l'action de l'insertion d'un message par LSB. Soit P la palette de pixels utilisée dans l'image, $P = \{0, 1, 2, ..., 255\}$ pour une image en nuances de gris sur 8 bits. Soit $\mathcal{P}$ un ensemble de paires de l'image (par exemple les paires de pixels adjacents, i.e un sous-ensemble de $P \times P$). On considère deux sous-ensembles X et Y de $\mathcal{P}$, chacun invariant après modification LSB. Le premier groupe est constitué de la réunion de trois ensembles A, B et C :

$$X \begin{cases} A & = \{(2n+1, 2n+2), (2n+2, 2n+1), n \in P\} \\ B & = \{(2n+1, 2n+3), (2n+3, 2n+1), (2n+2, 2n), (2n, 2n+2), n \in P\} \\ C & = \{(2n+3, 2n), (2n, 2n+3), n \in P\} \end{cases}$$

Le second groupe est constitué de la réunion de deux ensembles D et E :

$$Y \begin{cases} D &= \{(2n+1, 2n+1), (2n, 2n), n \in P\} \\ E &= \{(2n+1, 2n), (2n, 2n+1), n \in P\} \end{cases}$$

Le tableau suivant décrit le résultat de l'action d'une permutation LSB sur une paire d'un des ensembles précédents. L'opération 00 signifie qu'aucun élément de la paire n'est affecté par le LSB, 01 que le second élément de la paire est affecté, 10 le premier et 11 que les deux éléments de la paire sont modifiés.

Opération	D	E	A	B	C
00	D	E	A	B	C
01	E	D	B	A ou C	B
10	E	D	B	C ou A	B
11	D	E	C	B	A

Sous l'action d'une permutation LSB, chaque pixel de l'image a une probabilité p d'être modifié. On peut calculer sur l'image tatouée le nombre d'éléments de chacun des groupes précédents en fonction du nombre d'éléments des groupes initiaux avant permutation et de la probabilité p d'insertion. On considère les différentes paires possibles d'éléments des groupes précédents et on calcule pour chacune d'elles les différentes probabilité de changement d'état. On a résumé ces calculs dans le tableau 1.

Pour résoudre le système précédent, on fera l'hypothèse que $A = E$. Il n'y a en effet théoriquement aucune raison pour que le nombre de paires de type $(2n, 2n+1)$ soit foncièrement différent du nombre de paires de type $(2n+1, 2n+2)$.

$$\begin{cases} A' = & (A - B + C)p^2 + (B - 2A)p + A \\ B' = & -2(A - B + C)p^2 + 2(A - B + C)p + B \\ C' = & (A - B + C)p^2 + (B - 2C)p + C \\ D' = & 2(D - E)p^2 - 2(D - E)p + D \\ E' = & -2(D - E)p^2 + 2(D - E)p + E \end{cases}$$

On pose $a = 2(A' + B' + C') - 4(D' + E')$, $b = 4(-A' + D' + E') - 2B'$ et $c = 2A' - 2E'$, et on résoud le système précédent équivalent à l'équation du second degré $ax^2 + bx + c = 0$. La racine comprise entre 0 et 1 donne une évaluation de la longueur du message, en pourcentage de la taille du plan de bits.

E.3.2 Validité des hypothèses

Les deux hypothèses faites dans la stéganalyse du LIS proposée sont :

Groupe D		
Type de paire initiale	Type de paire finale	Probabilité associée
$(2n, 2n)$	$(2n+1, 2n+1)$,	D
$(2n, 2n)$	$(2n, 2n+1)$	$-p(1-p)D$
$(2n, 2n)$	$(2n+1, 2n)$	$-p(1-p)D$
$(2n, 2n+1)$	$(2n, 2n)$	$p(1-p)E$
$(2n, 2n+1)$	$(2n+1, 2n+1)$	$p(1-p)E$
$D' = 2(D-E)p^2 - 2(D-E)p + D$		

Groupe E		
$(2n, 2n+1)$	$(2n, 2n+1)$	E
$(2n, 2n)$	$(2n, 2n+1)$	$-p(1-p)E$
$(2n, 2n)$	$(2n+1, 2n)$	$-p(1-p)E$
$(2n, 2n+1)$	$(2n, 2n)$	$p(1-p)D$
$(2n, 2n+1)$	$(2n+1, 2n+1)$,	$p(1-p)D$
$E' = 2(D-E)p^2 - 2(D-E)p + E$		

Groupe A		
$(2n+1, 2n+2)$	$(2n+1, 2n+2)$	A
$(2n+1, 2n+2)$	$(2n+1, 2n+3)$	$-p(1-p)A$
$(2n+1, 2n+2)$	$(2n, 2n+2)$	$-p(1-p)A$
$(2n+1, 2n+2)$	$(2n, 2n+3)$	$-p^2A$
$(2n+1, 2n+3)$	$(2n+1, 2n+2)$	$p(1-p)B$
$(2n, 2n+3)$	$(2n+1, 2n+2)$	p^2C
$A' = (A-B+C)p^2 + (B-2A)p + A$		

Groupe B		
$(2n+1, 2n+3)$	$(2n+1, 2n+3)$	B
$(2n+1, 2n+3)$	$(2n+1, 2n+2)$	$-p(1-p)B$
$(2n, 2n+2)$	$(2n+1, 2n+2)$	$-p(1-p)B$
$(2n+1, 2n+2)$	$(2n+1, 2n+3)$	$p(1-p)A$
$(2n+1, 2n+2)$	$(2n, 2n+2)$	$p(1-p)A$
$(2n, 2n+3)$	$(2n, 2n+3)$	$p(1-p)C$
$(2n, 2n+3)$	$(2n, 2n+2)$	$p(1-p)C$
$B' = -2(A-B+C)p^2 + 2(A-B+C)p + B$		

Groupe C		
$(2n, 2n+3)$	$(2n, 2n+3)$	C
$(2n, 2n+3)$	$(2n+1, 2n+3)$	$-p(1-p)C$
$(2n, 2n+3)$	$(2n, 2n+2)$	$-p(1-p)C$
$(2n, 2n+3)$	$(2n+1, 2n+2)$	$-p^2C$
$(2n+1, 2n+3)$	$(2n, 2n+3)$	$p(1-p)B$
$(2n+1, 2n+3)$	$(2n, 2n+3)$	p^2A
$A' = (A-B+C)p^2 + (B-2A)p + A$		

TAB. 1 – Méthode LIS : calcul des groupes A, B, C, D et E

- Les images naturelles vérifient l'égalité statistique $|A| = |E|$
- Les modifications LSB sont équiprobables sur tous les groupes de pixels considérés. Autrement dit le schéma de tatouage ne s'adapte pas au contenu.

La deuxième hypothèse est bien vérifiée par notre algorithme de tatouage. Nous avons vérifié la première sur l'ensemble des images testées. Les images 512x512 donnent un rapport $\frac{|A|}{|E|}$ valant 0,941 en moyenne, avec un écart-type de 0.047. Les images 2560x1920 donnent un rapport $\frac{|A|}{|E|}$ valant 0,969 en moyenne, avec un écart-type de 0.119. Les mesures effectuées et les analyses ont donc été réalisées sous des hypothèses assez bonnes.

E.3.3 Résultats

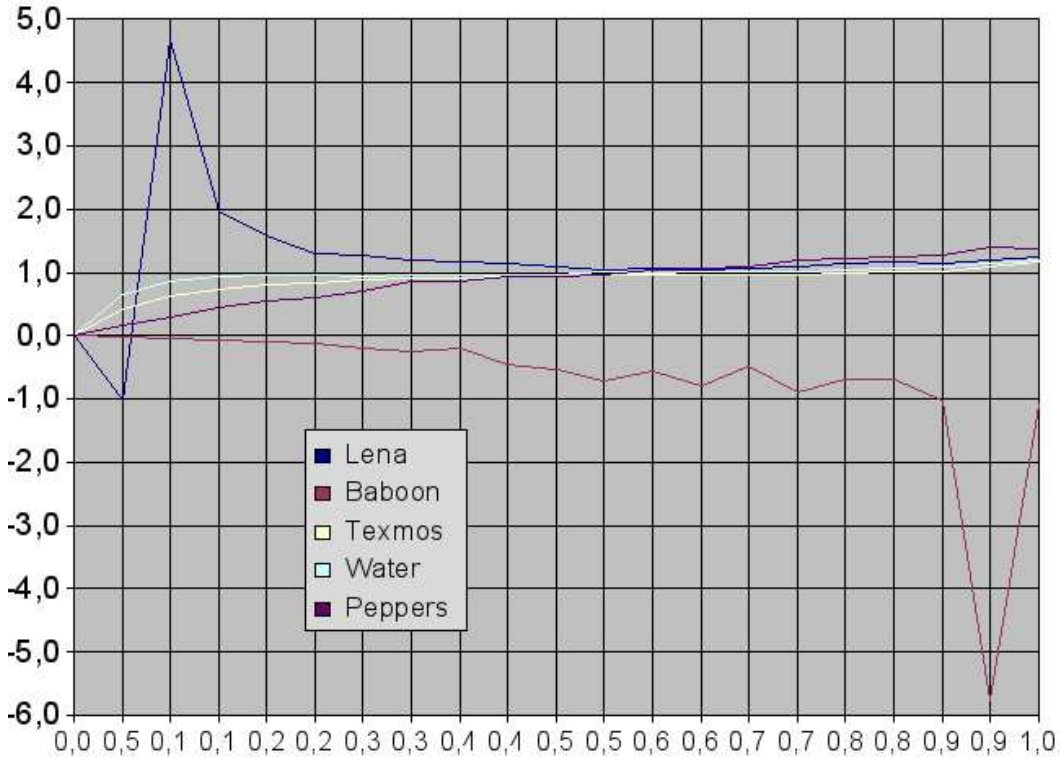

FIG. 7 – Courbes des taux d'erreur de détection de la méthode LIS sur les images petit format (512x512).

Le graphique en figure 7 synthétise les résultats de la stéganalyse LIS sur les images de taille 512x512. La stéganalyse est complètement inefficace sur Baboon, laissant présager de mauvais résultats sur des images très texturées. A l'exception de Baboon, on remarque que toutes les images réagissent bien dans la plage d'insertion [0,5-0,7]. Les meilleurs résultats sont obtenus sur TexMos. Ces premiers tests montrent que ces stéganalyses sont sensibles à des comportements statistiques d'ordre supérieur à un. Les variances

Bruit512	5408.235313	1071.899588
Lena	2290.151772	-9097.168811
Baboon512	1790.279936	-23277.817133
Water512	1885.906864	79550.292101
Peppers512	2902.702573	-37588.550760
Texmos512	5496.626128	2829.168370
Latronche	8408.139423	625269.012235
Belledonne	4263.374348	-232191.294094
Sappey	6461.519643	-127.233812
SolPinede	3789.317497	206197.227415
Futaie	2400.120331	-84629.584994
Ciel	336.932194	990.091604
Pelouse	1721.152457	7382.744687
Mur		-85186.464967

TAB. 2 – Variance et skewness des images

et skewness calculées pour ces images ne montrent en effet pas de corrélation avec les observations (cf. tableau 2) : TexMos et Peppers possèdent par exemple une variance importante mais les résultats sont très mauvais sur Peppers.

Pour étudier les erreurs sur les images Lena, Baboon et Peppers, nous avons donc étudié l'évolution des cardinaux des ensembles A, B, C, D et E en fonction de la taille du message inséré. La figure 8 présente les courbes d'évolution des cardinaux des ensembles considérés en fonction de la taille du message inséré. On voit que l'écart initial entre les cardinaux des ensembles A, C et D, E joue sur la précision de la détection. A l'aide du tableau de synthèse 3, on voit que les performances de la détection sont fortement atteinte quand cet écart est faible. Les images donnant de mauvais résultats possèdent des ratios $D-E$ et $A-C$ qui diffèrent d'un facteur 10 par rapport aux images donnant de bons résultats.

On observe des dégradations similaires sur les images grand format. Les images Pelouse et SolPinede possèdent toutes deux un écart entre les cardinaux des ensembles D & E, et A & C faible. Les performances sont moins bonnes sur ces images mais leur taille fait que le biais dans l'évaluation de la taille du message inséré est en général équivalent au biais de détection sur les bonnes images 512x512 (TexMos par exemple.). Les marquages sur Pelouse sont en effet détectés avec une bonne précision (5%) dans l'absolu, même si les biais pour les images moins texturées comme Ciel sont inférieur à 2%. On observe que pour deux images texturées Pelouse et SolPinede, les résultats sont biaisés de 5% pour Pelouse, ce qui rend la stéganalyse finalement assez efficace, alors que les biais sur SolPinede de plus de 10% même à 50% de taux d'insertion et le biais initial de 23% témoignent d'une image de grande capacité.

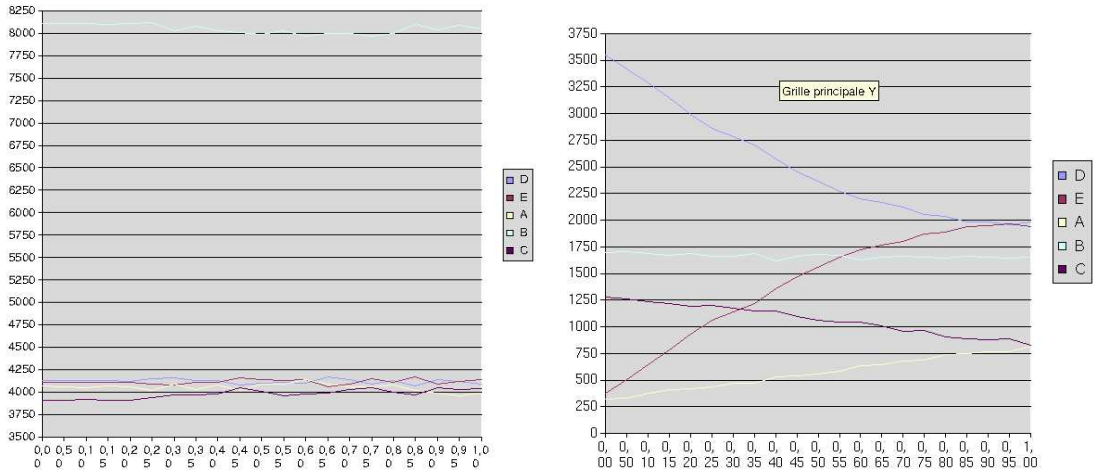

FIG. 8 – Diagrammes des groupes de pixels de l'analyse du LIS pour l'image Baboon à gauche, et TexMos à droite.

Image	D-E	A-C	(D-E)/(X+Y)	(A-C)/(X+Y)	A/E
Lena	1288	2897	0,02227	0,05010	1,01813
Baboon	68	107	0,00275	0,00433	0,98318
Water	1717	2763	0,03317	0,05338	0,98055
Peppers	184	1192	0,00372	0,02410	0,98153
Texmos	3601	-943	0,37440	0,09804	0,74230
Sappey	30825	23112	0,10242	0,07679	0,96070
Belledonne	249961	283838	0,13705	0,15563	1,00618
Latronche	265803	124128	0,21451	0,10017	0,89188
Ciel	405383	488574	0,17138	0,20655	0,99755
Pelouse	1766	4327	0,00460	0,01128	0,99426
Mur	311714	431474	0,13770	0,190609	1,00419
SolPinede	13889	18559	0,02367	0,03164	0,90034
Futaie	182532	235741	0,12162	0,15708	0,99806

TAB. 3 – Détail des cardinaux des ensembles X et Y de la méthode LIS.

FIG. 9 – Images à problèmes pour l'analyse du LIS : pelouse et solPinede

E.4 Memon

Nous étudierons dans cette partie une analyse stéganographique proposé par R. Chandramouli et N. Memon dans [8].

E.4.1 Principe

Dans cette méthode, on considère, sur l'ensemble des paires de pixels adjacents $\mathcal{P}$ deux sous-ensembles X et Y. La stéganalyse proposée par Memon est basée sur le fait que l'insertion d'un message dans le dernier plan de bit modifie l'égalité des cardinalités $|X| = |Y|$ que l'on suppose vraie pour les images naturelles. Cette hypothèse est équivalente à l'hypothèse selon laquelle le gradient des paires a une probabilité égale d'être positif ou négatif.

- $X = \{(u, v) \in \mathcal{P}$ tels que v soit pair et $u < v$, ou v impair, et $u > v\}$
- $Y = \{(u, v) \in \mathcal{P}$ tels que v soit pair et $u > v$, ou v impair, et $u < v\}$

On définit alors les sous-ensembles suivants :
- $Z \subset \mathcal{P}$ est l'ensemble des paires de la forme (u, u).
- $W \subset \mathcal{P}$ est l'ensemble des paires de la forme $(2k, 2k + 1)$ ou $(2k + 1, 2k)$
- $V \subset \mathcal{P}$ est le complémentaire de W dans Y : $V = Y - W$.

Alors $\mathcal{P} = X \cup W \cup V \cup Z$. On peut alors décrire le résultat d'une opération de type LSB sur une paire (u, v) d'un des ensembles précédents. On notera comme dans la section précédente 00 l'opération qui laisse inchangé u et v, 01 (resp. 10) l'opération qui modifie v (resp. u) par LSB et laisse u (resp. v) inchangé, et 11 l'opération qui modifie les deux éléments de la paire considérée. Le tableau suivant synthétise les résultats de l'action de ces opérations sur des paires des ensembles précédents.

Opération π	X	W	V	Z
00	X	W	V	Z
01	X	Z	V	W
10	V	Z	X	W
11	V	W	X	Z

On note $\rho(\pi, A)$ la probabilité qu'un pixel de $A \in \mathcal{P}$ soit modifié par l'opération π. La stéganalyse suppose que les modifications par LSB sur tout ensemble $A \in \{X, V, W, Z\}$ sont équiprobables. Autrement dit $\rho(\pi, A) = \rho(\pi, \mathcal{P})$ pour toute opération $\pi \in \{00, 01, 10, 11\}$ et tout ensemble $A \in \{X, V, W, Z\}$. Ainsi on a

$$\begin{aligned} \rho(00, \mathcal{P}) &= (1 - \tfrac{p}{2})^2 \\ \rho(01, \mathcal{P}) &= \rho(10, \mathcal{P}) = \tfrac{p}{2} \\ \rho(11, \mathcal{P}) &= (\tfrac{p}{2})^2 \end{aligned}$$

Si p est la taille du message, en pourcentage de la taille totale du plan de bits, alors la proportion de bits modifiés par l'insertion dans le plan de bits initial est égal à $p/2$. A l'aide des deux suppositions faites plus haut, on peut calculer les cardinaux des ensembles X', Y' et W', définis comme les ensembles X, Y et W mais sur l'image tatouée, en fonction des cardinaux des ensembles initiaux.

$$\begin{aligned} |X'| = &\quad |X|(1 - p/2) + |V|p/2 \\ |V'| = &\quad |V|(1 - p/2) + |X|p/2 \\ |W'| = &\quad |W|(1 - p + p^2/2) + |Z|p(1 - p/2) \end{aligned}$$

Alors on tire $|X'| - |V'| = (|X| - |V|)(1 - p)$. De plus, on a supposé que $|X| = |Y|$, donc $|X| = |V| + |W|$. Alors $|X'| - |V'| = |W|(1 - p)$. Si on pose $\gamma = |W| + |Z| = |W'| + |Z'|$ ($W \cup Z$ est invariant par permutation LSB) alors, en utilisant l'équation de $|W'|$ on trouve :

$$|W'| = (|X'| - |V'|)(1 - p) + \gamma p(1 - p/2)$$

D'autre part $|\mathcal{P}| = |X'| + |V'| + |W'| + |Z'| = |X'| + |V'| + \gamma$. L'équation précédente devient donc

$$1/2\gamma^2 p^2 + (2|X'| - \mathcal{P})p + |Y'| - |X'| = 0$$

Si $\gamma \neq 0$, la racine la plus petite de cette équation donne la valeur de p correspondante aux cardinaux des ensembles considérés dans l'image analysée. Dans le cas inverse, l'analyse échoue : les équations en p obtenues sont constantes.

E.4.2 Validité des hypothèses

Les trois hypothèses faites par Memon dans la stéganalyse proposée sont :
- Les images naturelles vérifient l'égalité statistique $|X| = |Y|$
- La constante $|W| + |Z|$ est non nulle .
- Les modifications LSB sont équiprobables sur tous les groupes de pixels considérés : autrement dit le schéma de tatouage ne s'adapte pas au contenu.

La dernière hypothèse est bien vérifiée par notre algorithme de tatouage. Nous avons vérifié les deux premières hypothèses sur l'ensemble des images testées. Les images 512x512 donnent un rapport $\frac{|X|}{|Y|}$ valant 0,9950 en moyenne, avec un écart-type de 0.0047. Les images 2560x1920 donnent un rapport $\frac{|X|}{|Y|}$ valant 0,9940 en moyenne, avec un écart-type de 0.0098. Toutes ces images ont en outre une valeur de γ non nulle. Les mesures effectuées et les analyses ont donc été réalisées sous les bonnes hypothèses.

E.4.3 Résultats

Le graphique de la figure 10 synthétise les résultats de la stéganalyse de Memon sur les images de taille 512x512. Les biais de détection calculés sur ces images sont généralement inférieurs à 5%. On constate que l'écart de détection est important pour des taux d'insertion élevés : au-delà de 95 %, la différence de détection dépasse les 5% pour atteindre les 10%. La précision est influencée par le biais initial. Les images présentant un biais initial élevé sont Baboon et TexMos. Malgré une amélioration de la détection avec la taille du message inséré, elles conserveront ces mauvais résultats initiaux, et la capacité de ces images s'en trouve donc augmentée.

On peut faire des observations similaires sur les images grands formats. Le biais initial sur les images Sappey, Latronche, Pelouse et SolPinede influence la précision de la détection à toutes les échelles, même si celle-ci s'accroît avec des taux d'insertion élevés.

Pour étudier le biais de détection, nous avons comparé les cardinaux des ensembles X, V, W et Z. Sur le tableau de synthèse 4, on constate que lorsque le cardinal des ensembles W et Z est faible comme pour Baboon et TexMos, l'erreur de détection est importante. On constate la même chose sur les images grand formats sur les images sus-mentionnées.

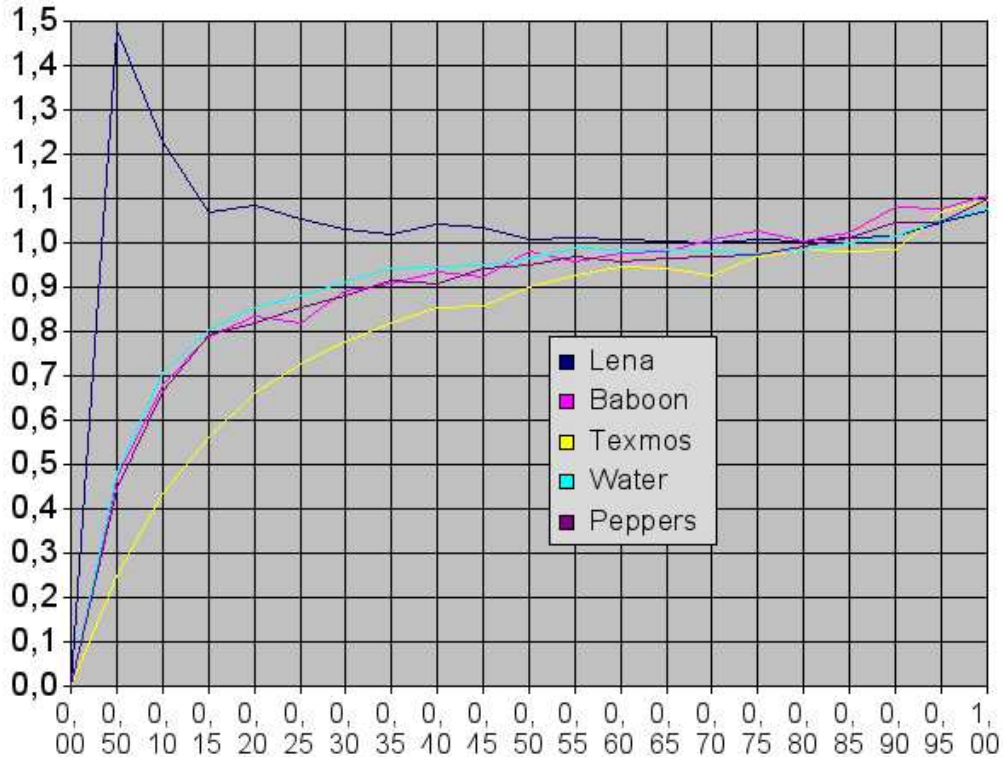

FIG. 10 – Courbes des taux d'erreur de détection de la méthode Memon sur les images petit format (512x512).

Image	X	V	W	Z	Y (V+W)
Lena	59872	49506	10203	11491	59709
Baboon	63316	59364	4162	4230	63526
Water	59766	50865	9362	11079	60227
Peppers	60890	52668	8665	8849	61333
TexMos	63057	62854	780	4381	63634
Sappey	261802	213007	54383	85208	267390
LaTronche	916458	537535	376823	626784	914358
Belledonne	971809	764700	227644	493447	992344
Ciel	759509	227272	532718	938101	759990
Mur	1195705	1129357	65386	67152	1194743
Pelouse	822660	321068	501079	812793	822147
SolPinede	1162059	1064056	108798	122687	1172854
Futaie	979940	664406	315361	497893	979767

TAB. 4 – Détail des cardinaux des ensembles X et Y de la méthode Memon.

Les erreurs de détection s'expliquent par des différences statistiques entre les groupes de pixels sur lesquels se basent les calculs de la stéganalyse. Celle-ci suppose en effet que le message inséré dans le dernier plan de bits se répartit également sur les différents ensembles défini par Memon. Or les groupes de cardinaux faibles présentent un écart par rapport à la prévision statistique importante, ce qui induit des erreurs d'évaluation de la taille du message en conséquence. La précision augmente avec l'augmentation de la taille, parce qu'alors l'écart entre les répartitions statistiques sur les différents groupes de paires de pixel tend à disparaître.

E.5 Fridrich

E.5.1 Principe

Nous décrivons dans cette section la méthode de stéganalyse RS proposée dans [16] dans le cas particulier des images en nuances de gris sur 8 bits. La stéganalyse consiste tout d'abord à construire une partition de l'image initiale comme réunion de groupes de pixels de taille n. Ces pixels prennent leur valeurs dans un ensemble P, qui vaut {0, 1, 2, ... , 255} pour une image en nuances de gris sur 8 bits. Si on note x_i le pixel d'indice i du groupe, on peut écrire le groupe $G = (x_1, ..., x_n)$.

On peut par exemple prendre comme nous l'avons fait dans nos expériences, quatre pixels consécutifs sur une ligne. Ensuite, on choisira une fonction f appelée fonction de discrimination, à valeur dans $\mathbb{R}$ et définie sur l'ensemble des groupes G. Cette fonction de discrimination reflète la texture ou rugosité du groupe de pixel G. On peut ainsi prendre pour f la fonction définie de la manière suivante : $f(G) = f(x_1, ..., x_n) = \sum_{i=1}^{n-1} |x_{i+1} - x_i|$

Enfin on définit les fonctions de permutation des nuances de gris de la façon suivante :

$F_1(x)$ est l'opération de permutation LSB d'un pixel de l'image : $F_1 : 0 \leftrightarrow 1, 2 \leftrightarrow 3,$ ... , $254 \leftrightarrow 255$.

F_{-1} est l'opération de permutation LSB translatée d'un : $F{-}1 : -1 \leftrightarrow 0, 1 \leftrightarrow 2,$ ... , $255 \leftrightarrow 256$. On peut remarquer que $F_{-1}(x) = F(x+1) - 1 \forall x$

F_0 est la fonction identité : $F_0(x) = x$.

On étendra les définitions précédentes aux groupes de pixels G en utilisant un masque $M = (m_j)$ de taille n à valeur dans {0, 1, -1} qui spécifie quelle opération doit être effectuée sur chaque élément du groupe. On définit $F_M(G) = (F_{m_1}(x_1), ..., F_{m_n}(x_n))$ Ainsi pour un groupe de taille 4, on pourra utiliser le masque $M = [-1\,0\,-1\,0]$; et alors pour $G = (x_1, x_2, x_3, x_4)$, $F_M(G) = (F_{-1}(x_1), F_0(x_2), F_0(x_3), F_{-1}(x_4))$.

Alors, l'ensemble des groupes G peut être partitionné en trois ensembles R, S et U définis de la manière suivante :

$G \in R_M \iff f(F(G)) > f(G)$, G est dit régulier

$G \in S_M \iff f(F(G)) < f(G)$, G est dit singulier

$G \in U_M \iff f(F(G)) = f(G)$, G est dit inchangé

Le masque $-M$ est défini à partir de $M = (x_j)$ par $-M = (-x_j)$ pour tout j.

On note R_M le pourcentage de groupes de type R calculé à l'aide du masque M, et S_M le pourcentage de groupes de type S. L'hypothèse sur laquelle se base l'analyse est le fait que sur des images naturelles R_M et R_{-M} sont égaux, ainsi que S_M et S_{-M}. Le bruitage du plan LSB par la stéganographie LSB modifie cette égalité et fait tendre la différence $R - S$ vers 0 à mesure que la taille du message inséré augmente. Après permutation de 50% du dernier plan de bits (la taille du message est alors égal à la taille du plan de bit), on a $R_M = S_M$ (1).

L'objet de la stéganalyse RS est l'extrapolation des courbes R_M, R_{-M}, S_M et S_{-M} en fonction de la taille du message p afin de calculer leurs intersections et d'en déduire la taille p. Les courbes R_{-M} et S_{-M} sont modélisées par des droites, alors que les courbes R_M et S_M le sont par des équations du second degré (cf. diagramme 11).

Soit donc à analyser une image contenant un message de longueur p, en pourcentage du nombre total de pixels. L'algorithme de stéganalyse consiste donc d'abord en le calcul du nombre de groupes $R_M(p/2)$, $S_M(p/2)$, $R_{-M}(p/2)$ et $S_{-M}(p/2)$ de l'image analysée. Puis, en permutant par LSB tous les pixels de l'image (i.e. application de F_1 à tous les pixels) on peut calculer les cardinaux et les pourcentages des groupes $R_M(1-p/2)$, $S_M(1-p/2)$, $R_{-M}(1-p/2)$ et $S_{-M}(1-p/2)$. En supposant que les courbes R_M et R_{-M} se coupent en la même abscisse que les courbes S_M et S_{-M}, et que $R_M(1/2) = S_M(1/2)$ (cf. (1)), on déduit les équations des paraboles et leur intersection.

On commence par réaliser le changement de variable suivant : $x = \frac{z-p/2}{1-p}$. On peut ainsi écrire les équation de R_{-M} et S_M dans ce nouveau repère :

$$R_{-M} = R_{-M}(1) - R_{-M}(0))x + R_{-M}(0)$$

$$S_{-M} = S_{-M}(1) - S_{-M}(0))x + S_{-M}(0)$$

On cherche à présent les équations de R_M et S_M.

$$R_M : ax^2 + bx + c$$

$$S_M : a'x^2 + b'x + c$$

Les valeurs calculées aux bornes 0 et 1 fournissent : $c = R_M(0)$ et $a+b+R_M(0) = R_M(1)$, $c' = S_M(0)$ et $a' + b' + S_M(0) = S_M(1)$. L'hypothèse (1) se traduit par $R_M(1/2) = S_M(1/2)$. On peut donc écrire :

$$R_M(1/2) = a/4 + (R_M(1) - R_M(0) - a)1/2 + R_M(0) = -a/4 + \frac{R_M(1) + R_M(0)}{2}$$

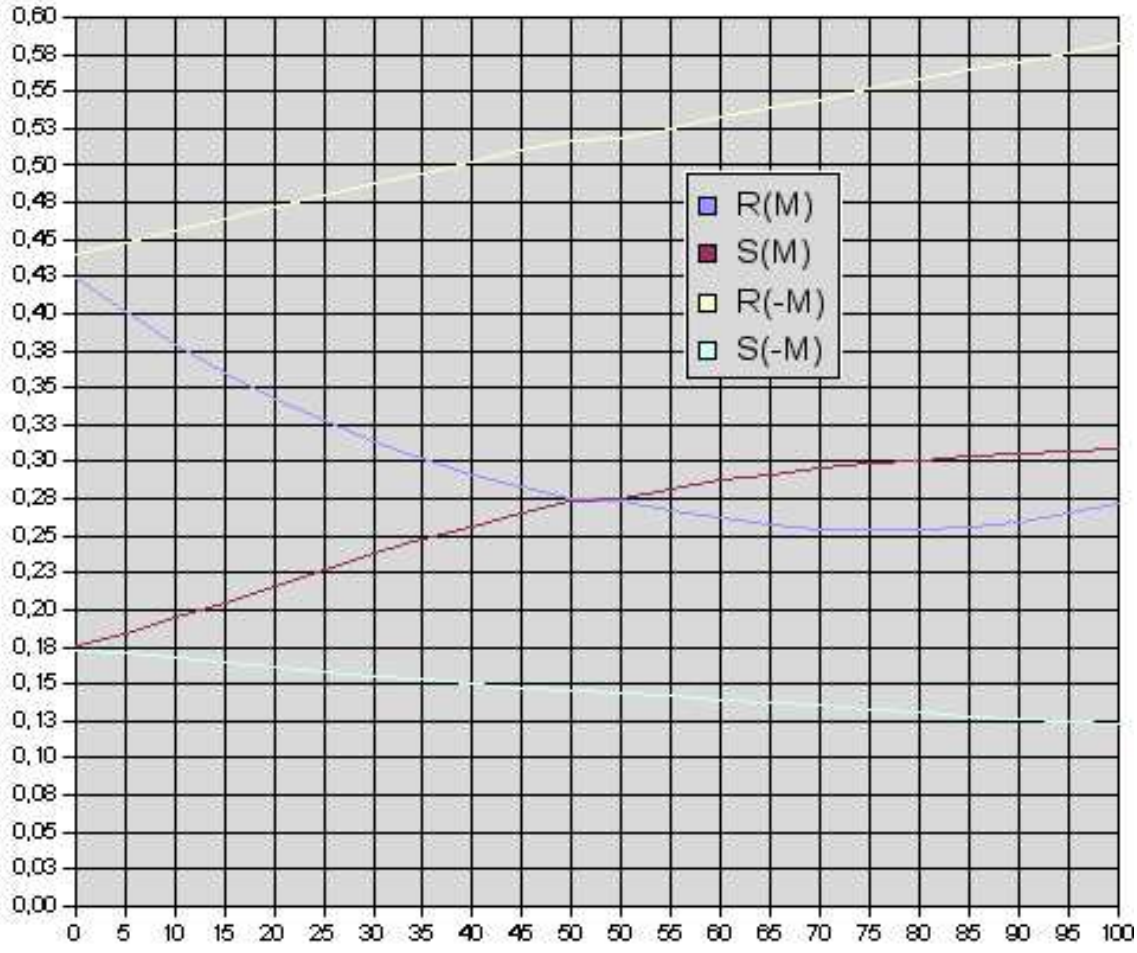

FIG. 11 – Diagramme RS de l'image Latronche

et

$$S_M(1/2) = a'/4 + (S_M(1) - S_M(0) - a')1/2 + S_M(0) = -a'/4 + \frac{S_M(1) + S_M(0)}{2}$$

De l'égalité de ces deux équations on déduit la suivante :

$$a = a' - 2\left(\underbrace{S_M(1) + S_M(0) - R_M(1) - R_M(0)}_{\alpha}\right)$$

On cherche donc l'intersection de R_{-M} avec R_M :

$R_{-M} \cap R_M : ax^2 + (R_M(1) - R_M(0) - a)x + R_M(0) = R_{-M}(1) - R_{-M}(0))x + R_{-M}(0)$

$R_{-M} \cap R_M : ax^2 + (R_M(1) - R_M(0) - R_{-M}(1) + R_{-M}(0) - a)x + R_M(0) - R_{-M}(0) = 0$

Et de même

$S_{-M} \cap S_M : a'x^2 + (S_M(1) - S_M(0) - S_{-M}(1) + S_{-M}(0) - a')x + S_M(0) - S_{-M}(0) = 0$

Ainsi en remplaçant a par $a' - 2\alpha$ dans la première équation et en égalisant les deux on obtient

$$-2\alpha x^2 + (R_M(1) - R_M(0) - R_{-M}(1) + R_{-M}(0) + 2\alpha)x + R_M(0) - R_{-M}(0) = (S_M(1) - S_M(0) - S_{-M}(1) + S_{-M}(0))x + S_M(0) - S_{-M}(0)$$

On pose ensuite $d_0 = R_M(0) - S_M(0)$, $d_1 = R_M(1) - S_M(1)$, $\alpha = -(d_0 + d_1)$, $d_{-0} = R_M(0) - S_{-M}(0)$, $d_{-1} = R_M(1) - S_{-M}(1)$. L'équation à résoudre devient

$$2(d_1 + d_0)x^2 + (d_{-0} - d_{-1} - 3d_0 - d_1)x + d_0 - d_{-0} = 0$$

La racine z de cette équation la plus petite fournit l'abscisse de $R_M \cap R_{-M}$ qui est également celle de $S_M \cap S_{-M}$. On retrouve la valeur de la longueur du message inséré p en procédant au changement de variable réciproque :

$$p = \frac{z}{z - 1/2}$$

E.5.2 Résultats

Le graphique de la figure 12 synthétise les résultats de la stéganalyse de Fridrich sur les images de taille 512x512 . On constate que la précision de détection est généralement inférieure à 5% pour toutes les images. Les résultats sont très biaisés sur l'image TexMos, où l'erreur de détection est quasiment constante et égale à 10% à tous les taux d'insertion inférieurs à 50%. Cette image possède donc une grande capacité pour cette méthode d'analyse. Les autres images ont un comportement plus régulier ; même si le biais initial de l'image influence la précision de la détection fortement en dessous de 15% pour la taille du message, cette précision et l'écart par rapport à la réalité diminuent rapidement au-delà de ce taux.

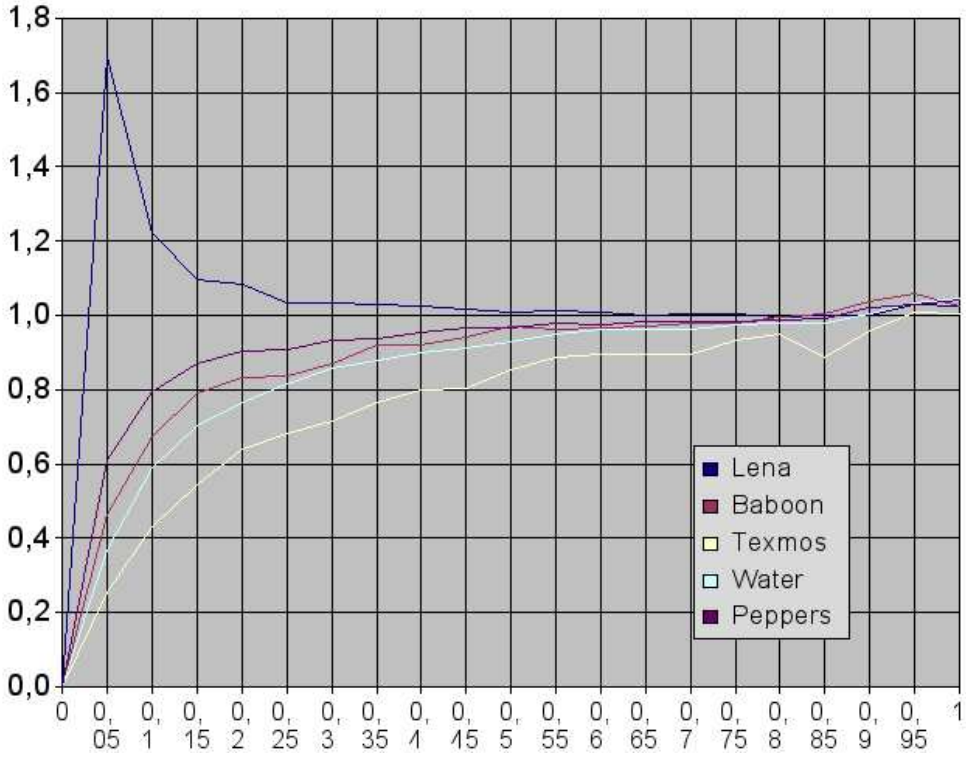

FIG. 12 – Courbes des taux d'erreur de détection de la méthode Fridrich sur les images petit format (512x512).

E.6 Comparaisons des méthodes

La méthode Fridrich est celle qui possède la plus grande stabilité sur tous les types d'image. La méthode de Memon est moins efficace qu'elle en général. Toutes deux achoppent sur l'image TexMos en présentant des résultats très biaisés (voir la figure 13). Sur ce type d'image, la méthode de Memon se révèle en fait moins mauvaise. La méthode du Lis est globalement moins précise que les deux précédente. Toutefois, cette méthode est très sensible à la texture des images. Baboon offre par exemple pour cette méthode une capacité maximale. A l'inverse la précision de détection sur TexMos supplante les deux autres. Sur toutes les méthodes, la précision de détection augmente du biais initial au taux de 85%, puis rechute. La précision de la détection est très bonnes sur les images peu texturées : les taux d'erreurs sur Ciel, Mur, Belledonne ou Futaie sont en effet inférieurs à 1% à tous les niveaux d'insertion. Ces images possèdent donc une capacité quasi nulle.

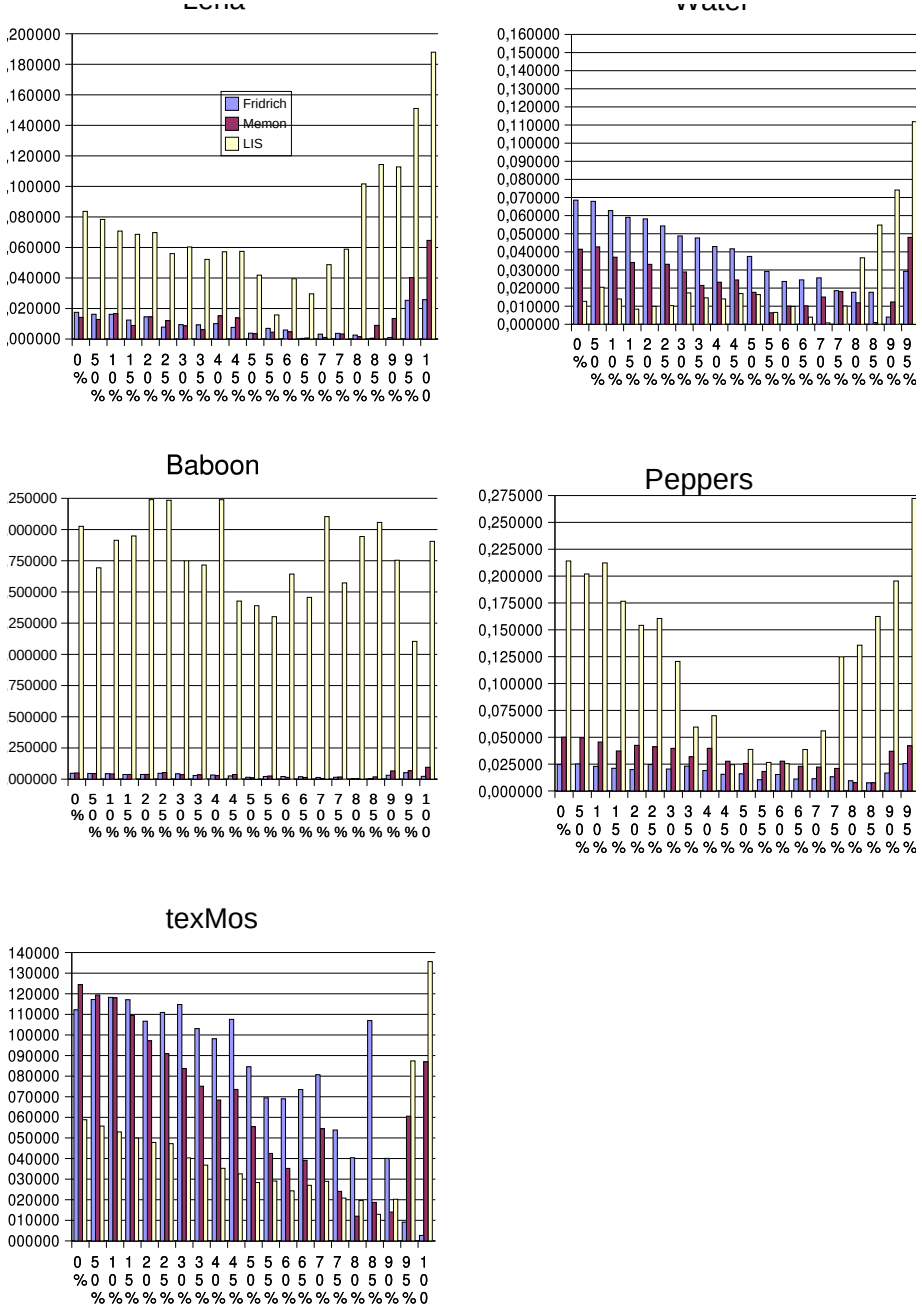

FIG. 13 – Comparaison des biais de détection des différentes méthodes sur les images 512x512

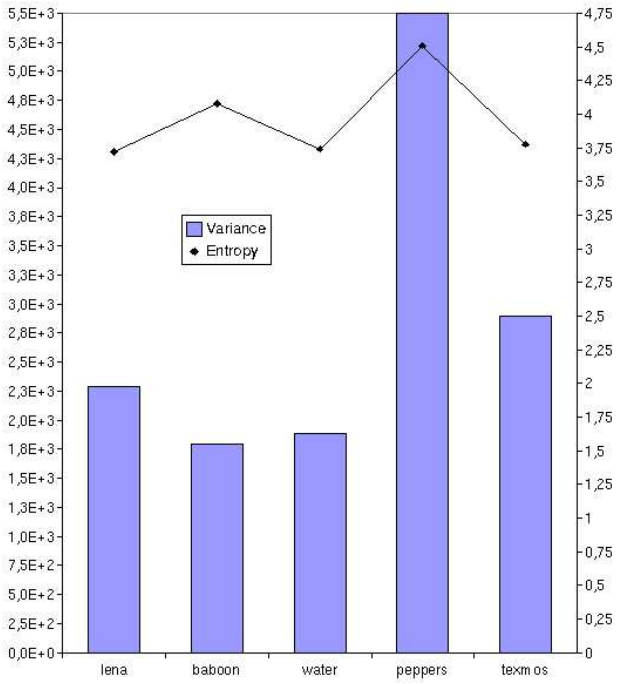

FIG. 14 – Diagrammes des variance et entropie pour les images petit format.

E.6.1 Matrice de cooccurrence

La matrice de cooccurrence d'une image permet de déterminer la fréquence d'apparition d'un motif formé de deux pixels séparés par une certaine distance d dans une direction particulière θ. Nous nous sommes ici focalisés sur un angle de $\theta = 0$: c'est en effet l'angle utilisé par les analyses présentées dans ce document. Cette matrice $MC = (MC_{i\,j})_{i,\,j\in\{1\,...,255\}}$ est définie pour tout couple de niveaux de gris (i, j) de la façon suivante :

$$MC_{i\,j} = card\{(p, p+1) \in Image^2 \mid I[p] = i,\ I[p+1] = j\ \}$$

avec p un pixel de l'image, $p+1$ le pixel adjacent à p et $I[p]$ la valeur du niveau de gris du pixel p. $MC_{i\,j}$ représente donc le nombre de couples de pixels $(p, p+1)$ de l'image tels que p et $p+1$ ont pour niveau de gris i et j respectivement.

L'entropie est définie de la façon suivante :

$$E = 1 - \frac{1}{N_c ln(N_c)} \sum_i \sum_j MC_{i\,j}.ln(MC_{i\,j})$$

où N_C est le nombre de couples $(p, p+1)$ entrant dans la composition de la matrice. Cette entropie est faible si on a souvent le même groupe de pixels, forte si chaque couple est peu représenté. Elle fournit un indicateur du désordre que peut présenter une texture.

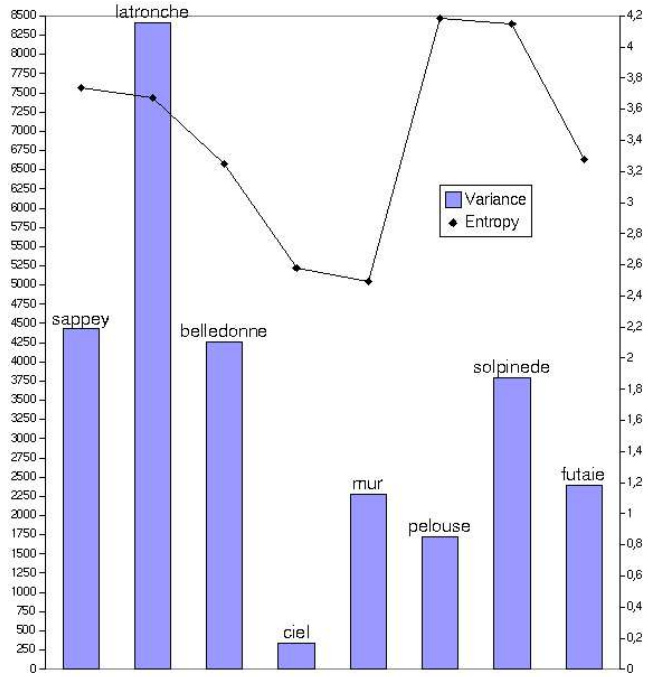

FIG. 15 – Diagrammes des variance et entropie pour les images grand format.

Nous avons calculé les variances et entropies de nos images de test en utilisant le programme pgmtexture sous Linux réalisant les calculs des matrices de cooccurrence. En examinant les diagrammes de synthèse construits à partir de ces valeurs et présentés en figure 14 et 15 on constate que les images ne réagissant pas bien à la stéganalyse du LIS ont toutes un entropie beaucoup plus forte que les images réagissant bien. Ceci se visualise également sur le diagramme 16 où sont représentées des vues aériennes des matrices de cooccurrences de Ciel et SolPinede : Ciel réagit bien à la stéganalyse et possède une entropie faible que l'on visualise par le fait que les valeurs de la matrice sont concentrées très proches sur une petite diagonale de la matrice. A l'inverse, SolPinede ne réagit pas bien à la stéganalyse et possède une entropie forte : la matrice possède des valeurs non nulles quasiment sur l'ensemble de la matrice.

E.7 Amélioration par filtrage Laplacien

Rappel

Le filtrage d'un image à l'aide d'un filtre Laplacien consiste à réaliser le produit de convolution de l'image avec un masque de moyenne nulle. L'image filtrée $\Lambda = |L \otimes X|$

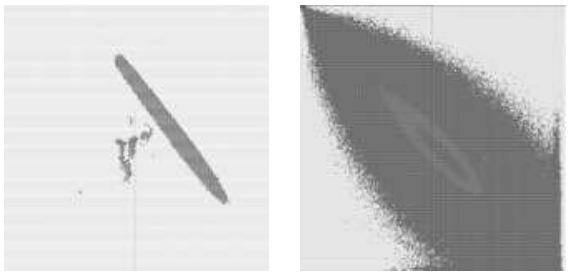

FIG. 16 – Vue aériennes des matrices de cooccurences des images Ciel et solPinede.

Image	Variance	Entropy
Lena	2290	3,719
Baboon	1792	4,076
Water	1887	3,742
Peppers	5496	4,509
Texmos	2898	3,777
Sappey	4430	3,737
Latronche	8410	3,674
Belledonne	4264	3,252
Ciel	337,8	2,577
Mur	2279	2,491
Pelouse	1722	4,188
Solpinede	3791	2401
Futaie	4,146	3,279

TAB. 5 – Matrices de cooccurence sur les images : variance et entropie

où X est l'image originale et L le masque

$$\begin{pmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{pmatrix}$$

pour un masque de type 4 et

$$\begin{pmatrix} -1 & -1 & -1 \\ -1 & 12 & -1 \\ -1 & -1 & -1 \end{pmatrix}$$

pour un masque de type 12 est ensuite seuillée pour ne retenir que certaines zones ; le filtrage par masque Laplacien met en évidence les zones texturées de l'image.

Au vu des résultats exposés dans la section précédente, nous avons cherché à améliorer les capacités de détection des stéganalyses en utilisant un filtrage de type Laplacien. L'objectif était de focaliser l'analyse sur les zones de l'images les moins dynamiques,

ayant constaté de meilleurs résultats sur des images à faible rugosité. Nous avons donc procédé en deux temps :

- filtrage de l'image initiale à l'aide d'un filtre Laplacien de type 4 ou 12
- création d'une image binaire à partir de l'image filtrée afin de ne retenir que les zones dépassant un certain niveau pour l'image filtrée (les images 17 représentent les masques obtenus).

L'analyse de l'image tatouée est alors réalisée seulement sur les zones identifiées dans l'image binaire. Les résultats obtenus n'ont pas été satisfaisants. Le filtrage est trop grossier et ne permet pas de relever d'amélioration de précision de la détection. En général, le biais de détection en est augmenté, sauf peut-être pour la méthode Memon sur Lena dans les taux d'insertion inférieurs à 30%, où l'écart de détection fut un peu resserré.

E.8 Amélioration des schémas d'insertion.

A partir de l'article de Memon [7], nous avons tenté d'améliorer le schéma d'insertion LSB afin de diminuer l'efficacité des méthodes d'analyse. A l'inverse du travail relaté dans le paragraphe précédent, l'idée de l'insertion fut de n'insérer un élément du message dans un pixel que si ce pixel était à distance supérieure à 3 de ses voisins directs (*).

Nous avons tout d'abord procédé en ignorant les lois de Kirchoff et en considérant que l'attaquant ignorait la méthode d'insertion employée. Les résultats sont alors sans équivoque. Les images passent au travers des stéganalyses à toutes les tailles de messages possibles, étant donné la contrainte (*). La méthode du LIS se base sur des ensembles de pixels éloignés de moins de 3 : l'insertion est donc complètement indécelable par cette méthode. Les deux autres attaques détectent des choses mais très éloignés de la réalité du taux d'insertion pratiqué. Sur Pelouse par exemple ou la capacité avec ce schéma atteint 25%, le taux calculé par la méthode Fridrich est de 10%. L'erreur est donc de 15% alors que pour un taux d'insertion de 25% avec le schéma LSB classique, la longueur estimée du message différait de 0,2% de la réalité. La méthode de Fridrich est donc plus efficace que le LIS sur les images texturées, mais nécessite quand même des zones d'insertion à faible rugosité pour bénéficier de la précision initiale. On notera que ce schéma d'insertion nécessiterait l'utilisation d'un code correcteur pour assurer la détection du message. En effet certaines paires de pixels $(2n, 2n + 3)$ sont affectées par l'insertion de manière à se trouver à distance 1. Il s'agit alors pour permettre la detection, d'ajouter de la redondance au message inséré, afin que les paires de pixels de l'image de type $(2n + 1, 2n + 2)$ soient écartées lors du décodage.

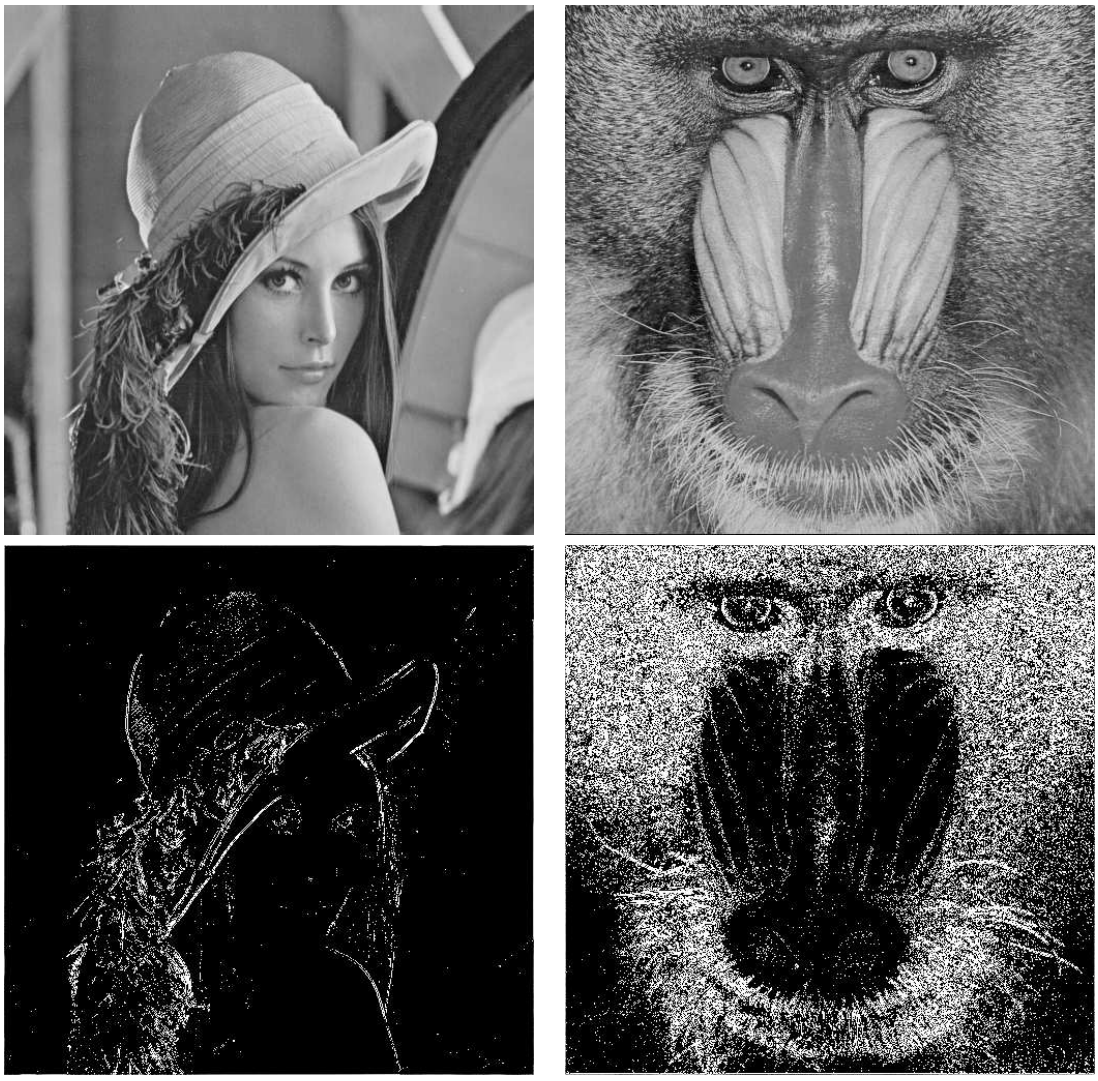

FIG. 17 – Images binaires de Lena et Baboon obtenues après filtrage et seuillage

F Stéganographie : les méthodes QIM

F.1 Description du schéma

Deux schémas d'insertion entrant dans la classe des schémas de quantification vectorielle ont été implémentés. Le premier insère un bit de message en certaines places de l'image déterminées aléatoirement. L'insertion $s(x, b)$ d'un bit b dans le bit x de l'image est alors réalisée comme décrit plus haut (C.2.1) :

$$\begin{cases} si\ b = 0 & s(x,b) = QF_\Delta(x) \\ si\ b = 1 & s(x,b) = QF_\Delta(x - \frac{\Delta}{2}) + \frac{\Delta}{2} \end{cases}$$

La seconde méthode implémentée utilise deux vecteurs de dilution $d[k, 0]$ et $d[k, 1]$. L'algorithme d'insertion peut s'écrire ainsi :

1. Diviser l'image en N blocs de taille L disjoints
2. Construire deux vecteurs $d[k,\ 0]$ et $d[k,\ 1]$ de dilution de taille L contenant les pas de quantification $\Delta_k \in [-\frac{\Delta}{2},\ \frac{\Delta}{2}]$. Ceux-ci sont construit avec la contrainte suivante :

$$d[k,\ 1] = \begin{cases} d[k,\ 0] + \frac{\Delta_k}{2},\ si\ d[k\,0) < 0 \\ d[k,\ 0] - \frac{\Delta_k}{2},\ si\ d[k\,0) \geq 0 \end{cases} ,\ k \in [1,\ L]$$

3. Le bloc i de l'image à tatouer est quantifié avec le quantificateur dilué $d[k,\ b_i]$ où b_i désigne le $i$ème bit à insérer.

Notre implémentation utilise les mêmes quantificateurs pour tous les blocs de longueur L. Cela n'est pas une nécéssité.

F.2 Résultats

L'étude approfondie de ce schéma d'insertion nous a fait penser que l'histogramme de l'image devait être perturbé par l'insertion d'un message par QIM. En effet, cet histogramme devrait, pour le premier algorithme d'insertion, comporter des pics à intervalles réguliers correspondant au fait que la quantification affecte tous les pixels porteurs d'un bit de message d'une même quantité $\pm\Delta$. C'est ce qu'on observe dans l'histogramme de Lena après insertion reproduit à la figure 18.

Pour la quantification par bloc avec vecteur de dilution, le même phénomène se produit, mais les pics sont plus nombreux : il existe autant de types de pics que de quantificateurs différents dans les vecteurs $d[k,\ 0]$ et $d[k,\ 1]$. Nous avons ainsi construit des histogrammes modulaires qui sont des histogrammes où les valeurs des nuances de gris sont vues modulo Δ. La figure 19 présente un tel histogramme pour l'image Lena. Les paramètres utilisés

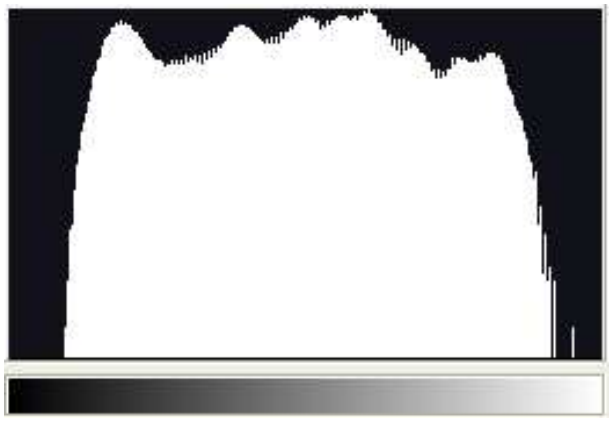

FIG. 18 – Histogramme de Lena après insertion QIM

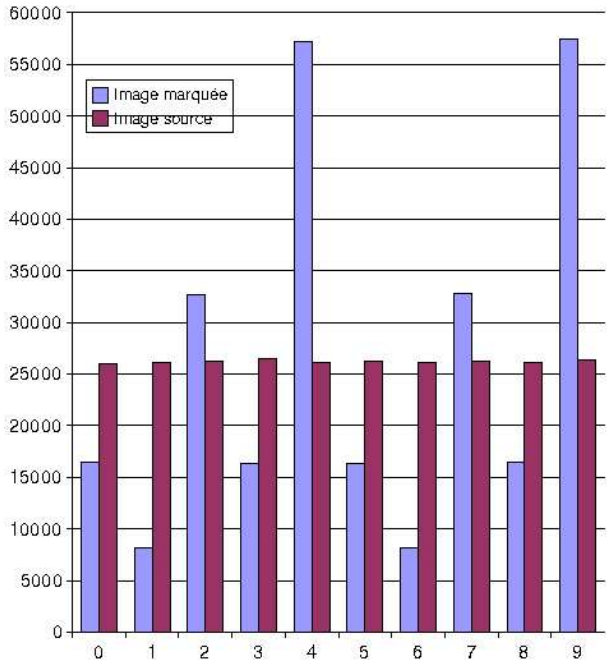

FIG. 19 – Histogramme modulaire de Lena après insertion QIM

Module	Nb quantificateurs
0	2
1	1
2	4
3	2
4	7
5	2
6	1
7	4
8	2
9	7

TAB. 6 –

pour l'insertion sont : une valeur de $\Delta = 10$ et une taille de bloc $L = 16$. A partir des valeurs des quantificateurs, on peut prévoir la répartition des pics de l'histogramme : pour un quantificateur Δ_k, l'histogramme modulaire présentera un pic pour la valeur $-\Delta_k \ mod \ \Delta$. Le tableau 6 donne les résultats de nos estimations à partir des vecteurs de dilution $d[k, 0]$ et $d[k, 1]$ utilisé pour l'insertion :

$$\begin{array}{lllllllllllllllll} d[k,0] = & (1 & 1 & 2 & 3 & 0 & -2 & 1 & 1 & 3 & 1 & 2 & -4 & 4 & 3 & 0 & 1) \\ d[k,0] = & (-4 & -4 & -3 & -2 & -5 & 3 & -4 & -4 & -2 & -4 & -3 & 1 & -1 & -2 & -5 & -4) \end{array}$$

On retrouve bien les résultats de l'histogramme modulaire.

Sachant que dans ce schéma, la taille du bloc L influence la taille du message inséré, il nous suffit d'estimer la taille du bloc pour réaliser une attaque. Cette première analyse du schéma QIM nous a conduit à imaginer une attaque possible du schéma, dont l'implémentation n'a pu être réalisée faute de temps. On parcourt l'image à l'aide d'une fenêtre égale à un multiple d'une taille bloc supposée initiale. Cette fenêtre est analysée bloc par bloc : on tente de détecter un module constant entre deux éléments de même indice sur des blocs différents. En effet, d'un bloc à l'autre les éléments de même indice sont quantifiés avec le même quantificateur scalaire.

Références

[1] T. Aura. *Practical invisibility in digital communication.* Springer- Verlag, 1996.

[2] I. Avcibas, N. Memon, and B. Sankur. Steganalysis using image quality metrics. In *Security and Watermarking of Multimedia Contents, SPIE. San Jose, CA*, February 2001.

[3] I. Avcibas, N. Memon, and B. Sankur. Image steganalysis with binary similarity measures. In *IEEE International Conference on Image Processing, Rochester, New York*, September 2002.

[4] P. Bas. *Méthodes de tatouages d'images fondées sur le contenu.* PhD thesis, Thèse de l'Institut National Polytechnique de Grenoble, France, 2000. manuscrit avaible on http ://www.lis.inpg.fr/pages_perso/bas/index.htm.

[5] C. Cachin. An information-theoretic model for steganography, 1998.

[6] R. Chandramouli. A mathematical approach to steganalysis.

[7] R. Chandramouli, G. Li, and N. Memon. Adaptive steganography. In *Security and Watermarking of Multimedia Contents IV*, 2002.

[8] R. Chandramouli and N. Memon. Analysis of lsb based image steganography techniques. In *Proceedings of the International Conference on Image Processing, Thessaloniki, Greece*, October 2001.

[9] B. Chen and G. W. Wornell. Implementations of quantization index modulation methods for digital watermarking and information embedding of multimedia. In *J. VLSI Signal Processing Syst. Signal, Image, and Video Technol. (Special Issue on Multimedia Signal Processing), vol. 27*, pages 7–33, Febrary 2001.

[10] Brian Chen and Gregory W. Wornell. Quantization index modulation : a class of provably good methods for digital watermarking and information embedding. In *IEEE Transaction on information theory, Vol. 47, N. 4*, pages 1423–1443, may 2001.

[11] H. Farid and L. Siwei. Detecting hidden messages using higher-order statistics and support vector machines. In *Pre-proceedings 5th Information Hiding Workshop, Noordwijkerhout, Netherlands*, October 2002.

[12] J. Fridrich and R. Du. Secure steganographic methods for palette images. In *The 3rd Information Hiding Workshop, LNCS vol. 1768, Springer-Verlag, New York*, 2000.

[13] J. Fridrich, R. Du, and M. Long. Steganalysis of lsb encoding in color images, 2000.

[14] J. Fridrich and M. Goljan. Practical steganalysis of digital images : State of the art. In *SPIE Photonics West, Vol. 4675, Electronic Imaging 2002, Security and Watermarking of Multimedia Contents, San Jose, California*, pages 1–13, January 2002.

[15] J. Fridrich and M. Goljan. Digital image steganography using stochastic modulation. In *EI SPIE Santa Clara, CA*, January 2003.

[16] J. Fridrich, M. Goljan, and R. Du. Reliable detection of lsb steganography in color and grayscale images. In *ACM Workshop on Multimedia and Security*, pages 27–30, 2001.

[17] J. Fridrich, M. Goljan, and D. Soukal. Higher-order statistical steganalysis of palette images. In *EI SPIE Santa Clara, CA, Jan 2003*, 2003.

[18] Jessica Fridrich, Miroslav Goljan, and Dorin Hogea. Attacking the outguess. In *ACM Workshop on Multimedia and Security 2002, Juan-les-Pins, France*, December 2002.

[19] J. J. Harmsen and W. A. Pearlman. Steganalysis of additive noise modelable information hiding. In *SPIE/IS&T Electronic Imaging SPIE Vol. 5022*, 2003.

[20] M. Goljan J. Fridrich and D. Hogea. Steganalysis of jpeg images : Breaking the f5 algorithm. In *5th Information Hiding Workshop, Noordwijkerhout, The Netherlands*, October 2002.

[21] M. Goljan J. Fridrich and D. Hogea. New methodology for breaking steganographic techniques for jpegs. In *EI SPIE Santa Clara, CA*, January 2003.

[22] N. F. Johnson and S. Jajodia. Steganography : Seeing the unseen. In *IEEE Computer*, pages 26–34, February 1998.

[23] N. F. Johnson and S. Katzenbeisser. *Information hiding - a survey.* Artech House, Norwood, MA, 2000.

[24] Neil F. Johnson and Sushil Jajodia. Steganalysis of images created using current steganography software. In *Information Hiding, Second International Workshop*, pages 273–289, 1998.

[25] Richard E. Newman, Ira S. Moskowit, LiWu Chang, and Murali M. Brahmadesam. A steganographic embedding undetectable by jpeg compatibility steganalysis. In *Lectures Notes in Computer Science - Information Hiding*, pages 258–277, 2003.

[26] F. Petitcolas, R. Anderson, and M. Kuhn. Information hiding - a survey. In *IEEE, 87(7)*, pages 1062–1078, July 1999.

[27] N. Provos and Peter Honeyman. Detecting steganographic content on the internet. Technical Report 01-1, University of Michigan, August 2001.

[28] Niels Provos. Defending against statistical steganalysis. In *10th USENIX Security Symposium*, pages 323–336, August 2001.

[29] Andreas Westfeld. High capacity depsite better steganalysis : F5- a steganographic algorithm. In *Fourth Information Hiding Workshop*, pages 301–315, 2001.

[30] Andreas Westfeld and Andreas Pfitzmann. Attacks on steganographic systems. In *3rd Info. Hiding Workshop, Dresden, Germany, September 28-October 1, LNCS vol. 1768, Springer-Verlag, New York*, 1999.