Stéganographie naturelle pour images JPEG

Théo Taburet, Patrick Bas, Wadih Sawaya, Jessica Fridrich

Stéganographie naturelle pour images JPEG

Théo TABURET[×], Patrick BAS[×], Wadih SAWAYA[♯], Jessica FRIDRICH[+]

[×]CNRS, Ecole Centrale de Lille
CRIStAL Lab, 59651 Villeneuve d'Ascq Cedex, France

[♯]IMT Lille-Douai, Univ. Lille, CNRS, Centrale Lille, UMR 9189, France

[+]Department of ECE
SUNY Binghamton, NY, USA

{theo.taburet, patrick.bas}@centralelille.fr, wadih.sawaya@imt-lille-douai.fr
fridrich@binghamton.edu

Résumé – Ce papier traite du développement d'un algorithme permettant d'étendre la "Stéganographie naturelle" (NS) aux images JPEG. Cette méthode permet l'insertion un message de taille importante avec une sécurité empirique élevée. Elle utilise le principe de changement de source : partant d'une image cover capturée à ISO_1 le signal stéganographique est généré de sorte à imiter une image capturée à $ISO_2 > ISO_1$. Au niveau du capteur, ce changement de sensibilité peut être modélisé par l'ajout d'un bruit photonique indépendant suivant une loi Normale. La problématique de cette méthode réside ici dans la modélisation de ce signal après développement dans le domaine DCT. Partant d'une image RAW nous prenons en compte les étapes du développement (le dématriçage, la transformation luminance et la quantification JPEG) afin de modéliser les dépendances entre les coefficients DCT voisins via le calcul d'une matrice de covariance. La simplicité de ce schéma de développement permet de justifier son intérêt qui offre des taux d'insertions supérieurs à l'état de l'art ($>$ 1 bpnzAC) tout en préservant la sécurité pratique ($P_E \simeq 40\%$).

Abstract – This paper discusses the development of an algorithm to extend "Natural Steganography" (NS) to JPEG images, this method could allow an insertion of a large size message with high empirical security. This schema uses the principle of cover source switching: starting from a cover image captured at ISO_1 the steganographic signal is generated to imitate an image captured at $ISO_2 > ISO_1$. At the sensor level, this change in sensitivity can be modelled by adding an independent photonic noise according to a Normal distribution. The problem of this method lies here in the modeling of this signal in the DCT domain. Starting from a RAW image, the processing steps considered are demosaicking (bi-linear), luminance averaging and quantification related to JPEG compression. They are modeled using a covariance matrix. The simplicity of this development scheme justifies the interest of this scheme which offers superior embedding rates compared to the state of the art ($>$ 1 bit per non-zero coefficient) while preserving high practical security ($P_E \simeq 40\%$).

1 Stéganographie naturelle

La stéganographie se présente comme l'art de dissimuler de l'information dans un média anodin appelé "cover", permettant la transmission de messages secrets entre des pairs. Le produit de cette opération de dissimulation s'appelle "stégo". L'insertion d'un message induit une distorsion du modèle initial qui va potentiellement permettre au stéganalyste de distinguer la stégo d'une cover. Le but de la stéganographie naturelle (NS) est de tâcher de faire passer l'insertion pour un processus naturel afin d'imiter le modèle d'une source cover.

1.1 Principes de base

On cherche ici à modéliser le signal stégo comme étant un signal additif permettant de produire une stégo (S_2) ayant les même propriétés qu'une cover capturée à $ISO_2 > ISO_1$ et ce à partir d'une image cover (S_1) capturée à ISO_1 : il s'agit du principe du "changement de source cover" permettant de passer d'un modèle de source $\mathcal{S}_1$ à une source $\mathcal{S}_2$ [1]. Cette variation des propriétés statistiques est due au changement de sensibilité (la scène capturée étant identique). En effet, lors de l'acquisition d'une image les erreurs de comptage des photons sur les photosites induisent la génération d'un bruit qui s'exprime indépendamment sur chaque photosite. Pour une sensibilité ISO_1 donnée, le bruit photonique sur le photosite (i, j) peut être modélisé par une loi de Poisson et approximé par une loi Normale de moyenne nulle et dont la variance est en relation affine avec la valeur non bruitée du photosite ($\mu_{i,j}$) :

$$N_{i,j}^{(1)} \sim \mathcal{N}(0, a_1\mu_{i,j} + b_1)$$

La somme de variables normales indépendantes restant une variable normale, et en supposant que $\mu_{i,j} \simeq x_{i,j}$, le signal stéganographique S sera alors distribué selon :

$$S_{i,j} \sim \mathcal{N}(0, (a_2 - a_1)x_{i,j}^{(1)} + (b_2 - b_1)\ )$$

Les propriétés statistiques d'une stégo étant ainsi les mêmes que celles d'une cover capturée à ISO_2 nous nous proposons

de déterminer le niveau de sécurité empirique, a priori élevé de cette approche.

1.2 Insertion dans le domaine JPEG

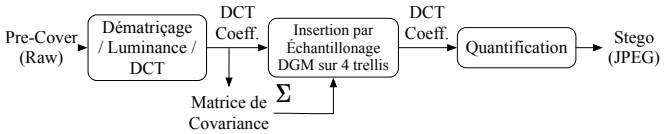


FIGURE 1 – Processus d'insertion présenté dans ce document.

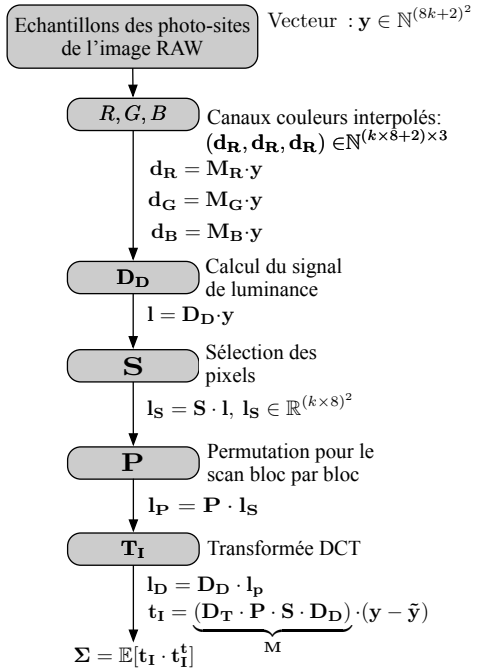


FIGURE 2 – Les différentes étapes permettant le calcul de la matrice de covariance Σ.

Dans cet article, nous étudions la NS lorsque la pré-cover est acquise à l'aide d'un capteur de couleur, développée par dématriçage linéaire, convertie en niveaux de gris, et compressée en JPEG. L'insertion est réalisée directement dans le domaine JPEG sur les coefficients DCT. La différence entre les capteurs monochrome et couleur a été étudiée avec la conclusion qu'une insertion sur les coefficients DCT offre une sécurité empirique élevée pour les capteurs monochromes, mais pas pour les capteurs couleurs [2]. Ceci est dû au fait que le dématriçage introduit des dépendances entre les coefficients DCT voisins. Lorsque ces dépendances ne sont pas prises en compte, l'insertion devient hautement détectable. Pour préserver les dépendances, le schéma d'insertion présenté dans cette étude utilise les matrices de covariance calculées à partir des observations de la pré-cover et de son développement, afin de générer les modifications sur l'image stégo.

Le schéma d'insertion est résumé dans la Figure 1 et peut être décomposé en différentes étapes décrites ci-contre :

1. Calcul de la matrice de covariance entre les coefficients DCT de $k \times k$ blocs voisins de 8×8 coefficients chacun. Cette partie est détaillée dans la section suivante.
2. Développement (dématriçage, luminance, et DCT). L'image pré-cover RAW est développée pour générer une image JPEG en niveaux de gris.
3. Échantillonnage suivant quatre treillis. Les auteurs ont montrés dans [8] que les dépendances intra et inter blocs sont dues au dématriçage et que les blocs DCT ne se touchant pas sont indépendants. Par conséquent les auteurs ont développé un schéma d'insertion permettant d'utiliser quatre treillis $\{\Lambda_1, \Lambda_2, \Lambda_3, \Lambda_4\}$. Les coefficients appartenant à Λ_1 sont échantillonnés indépendamment, les autres sont échantillonnés conditionnellement à leur voisinage (voir Figure 3a). Il est possible de montrer que pour un développement linéaire, la distribution du signal stégo pour chaque treillis $\Lambda_j (j \in \{1,..,4\})$ suit une distribution gaussienne multivariée (DGM) :
$$\mathcal{N}(\mathbf{m}_{i,j}, \mathbf{\Sigma}_{i,j}),$$
où pour $j \in \{2,3,4\}$ le vecteur d'espérance $\mathbf{m_{i,j}}$ et la matrice de covariance $\mathbf{\Sigma}$ de la distribution conditionnelle sont calculés en utilisant le complément de Schur de la matrice de covariance calculée.
4. Quantification JPEG. Le signal stégo simulé est quantifié à l'aide de la matrice de quantification JPEG pour un facteur de qualité (QF) donné.

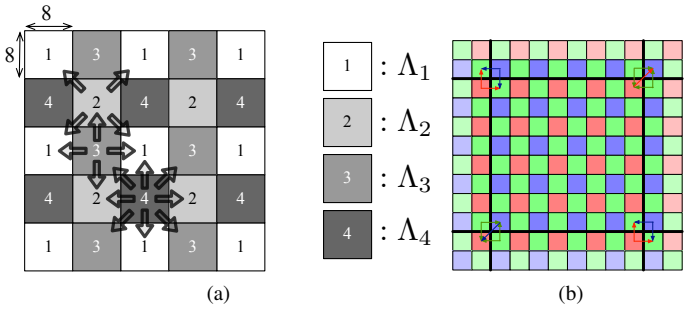


FIGURE 3 – (a) Les quatre treillis utilisés pour l'insertion. (b) Emplacements des photosites (couleurs foncées) utilisés pour prédire les valeurs des pixels dans un bloc (dématriçage bi-linéaire). Les blocs diagonaux sont impliqués dans le calcul sur deux pixels pour le canal bleu (en haut à droite) et le canal rouge (en bas à gauche).

2 Calcul de la matrice de covariance

Nous considérons ici le dématriçage le plus simple (par soucis calculatoire) : l'interpolation bi-linéaire. Pour calculer la matrice de covariance $\mathbf{\Sigma}$ du signal stégo dans le domaine DCT, nous devons décomposer le processus de développement en une séquence d'étapes linéaires résumées ci-dessous :

1. Sélection de $k \times k$ blocs de photosites, en incluant la bordure externe permettant d'avoir les photosites nécessaires à l'interpolation de ceux de la bordure interne (Figure 3b).
2. Dématriçage.
3. Calcul du signal de luminance.
4. Recadrage sur $k \times k$ blocs, avec $k \in \{1, 3\}$.
5. Permutation pour passer d'un balayage ligne par ligne à bloc par bloc.
6. Transformation DCT-2D.
7. Calcul de $\mathbf{\Sigma}$.

Ces opérations (Figure 2) peuvent ainsi s'écrire (voir [3]) comme une multiplication matricielle $\mathbf{t_I} = \mathbf{M} \cdot (\mathbf{y} - \tilde{\mathbf{y}})$ où $\mathbf{y} \in \mathbb{N}^{(8k+2)^2}$ est un vecteur d'observations obtenu à partir de $k \times k$ blocs de photosites et $\mathbf{t_I} \in \mathbb{N}^{(8k)^2}$ le vecteur correspondant aux blocs DCT avant quantification JPEG.

Enfin $\mathbf{\Sigma} = \mathbb{E}[\mathbf{t_I} \cdot \mathbf{t_I^t}]$.

3 Analyse de la matrice de covariance

La matrice de covariance, calculée sur 2×2 blocs adjacents, est illustrée pour le cas d'un capteur couleur (Figure 4). On peut noter que pour un capteur monochrome $\mathbf{\Sigma_{mono}}$ aurait été diagonale contrairement à $\mathbf{\Sigma}$ qui affiche de nombreuses corrélations (intra et inter-blocs). On peut subdiviser cette matrice en quatre types de matrices $\in \mathbb{R}^{64 \times 64}$ (voir Figures 4 et 5).

Situées sur la diagonale, les $\mathbf{\Sigma_I}$ capturent les corrélations entre les coefficients DCT du même bloc (intra-correlations).

Les corrélations inter-blocs sont capturées par $(\mathbf{\Sigma}_{\rightarrow}, \mathbf{\Sigma}_{\leftarrow})$ pour les blocs horizontaux, $(\mathbf{\Sigma}_{\uparrow}, \mathbf{\Sigma}_{\downarrow})$ pour les blocs verticaux, et $(\mathbf{\Sigma}_{\nearrow}, \mathbf{\Sigma}_{\swarrow}, \mathbf{\Sigma}_{\searrow}, \mathbf{\Sigma}_{\nwarrow})$ pour les blocs diagonaux.

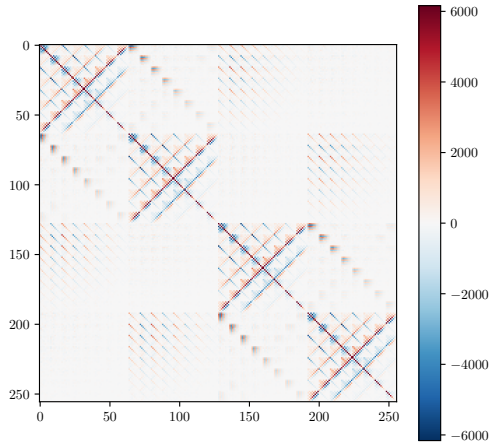


FIGURE 4 – Matrice de covariance $2 \times 2 \times 8 \times 8$ des coefficients DCT pour un capteur couleur. Ici, la dynamique des corrélations a été seuillée à $[-6000, 6000]$ pour des raisons d'affichage.

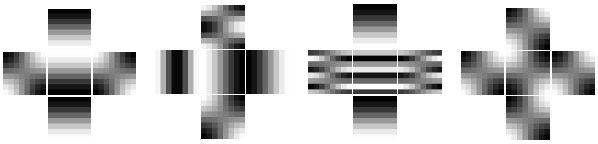
Les origines de ces corrélations résident dans l'opération de dématriçage qui induit inévitablement des corrélations avec les coefficients voisins. A partir de l'analyse de cette matrice de covariance et de son calcul analytique, nous pouvons tirer les conclusions suivantes :

- Les corrélations les plus importantes correspondent aux blocs connectés horizontalement et verticalement car pour prédire les pixels de la bordure interne d'un bloc de nombreux photosites des blocs voisins horizontaux et verticaux sont nécessaires.
- Les corrélations avec les blocs voisins diagonaux sont plus faibles car la prédiction des pixels situés aux coins ne fait ainsi intervenir qu'un unique photosite (voir Figure 3b).

Σ_I	$\Sigma_{\rightarrow}$	$\Sigma_{\downarrow}$	$\Sigma_{\searrow}$
$\Sigma^t_{\rightarrow}$	Σ_I	$\Sigma_{\swarrow}$	$\Sigma_{\downarrow}$
$\Sigma^t_{\downarrow}$	$\Sigma^t_{\swarrow}$	Σ_I	$\Sigma_{\rightarrow}$
$\Sigma^t_{\searrow}$	$\Sigma^t_{\downarrow}$	$\Sigma^t_{\rightarrow}$	Σ_I

FIGURE 5 – Subdivisions de la matrices de covariance représentant les 9 sous-matrices détaillées précédemment.

- Les blocs en 4-connexité sont inter-corrélés via une variable aléatoire partagée, ces corrélations sont très faibles.



(a) (1,0) (b) $(0, 1)$ (c) $(0, 5)$ (d) $(1, 1)$

FIGURE 6 – Différents agencements des modes les plus corrélés pour un bloc central et son voisinage en 4-connexité.

De cette matrice de covariance on peut dresser des conclusions sur la consistance fréquentielle et la continuité spatiale (voir Figure 6), qui rappellent des travaux similaires utilisant des insertions synchrones [4].

4 Résultats

La sécurité empirique est évaluée comme

$$P_{\mathrm{E}} = \min_{P_{\mathrm{FA}}} \frac{1}{2}(P_{\mathrm{FA}} + P_{\mathrm{MD}}),$$

avec P_{FA} et P_{MD} représentant le taux de fausses alarmes et de détections manquées.

La stéganalyse a été effectuée avec le jeu de caractéristiques DCTR [5] et un classifieur à faible complexité [6].

Nous évaluons la sécurité empirique de la NS dans le domaine JPEG pour 200 images acquises avec un capteur de couleur (Z-CAM-E1[X]) pour un changement de source de $ISO_1 = 100$ à $ISO_2 = 200$. Les images RAW ont été subdivisées en images 512×512 pour obtenir 9600 images.

4.1 Performance par rapport à l'état de l'art :

Nous comparons les schémas d'insertion suivants :

- La pseudo-insertion, l'insertion est simulée au niveau des photosites. Ces résultats peuvent être considérés comme un objectif de référence mais ne reflètent pas une insertion pratique dans le domaine JPEG.
- 4-treillis : Le schéma utilise la matrice de covariance calculée analytiquement pour réaliser l'échantillonnage.
- SI-UNIWARD [7] : Afin de réaliser une comparaison avec l'état de l'art le taux d'insertion a été fixé à 1 bpnzAC qui correspond à la capacité d'insertion maximale de ce schéma.

En utilisant 4-treillis, la capacité d'insertion varie d'environ 1 bpnzAC à QF 75 à 2 bpnzAC à QF 100. Tout d'abord, on remarque dans le tableau 1 qu'il n'y a pas de différence notable de sécurité empirique entre une pseudo-insertion sur les photosites et une insertion dans le domaine JPEG avec 4-treillis, ce qui valide la pertinence de notre approche dans sa capacité à générer des modification en préservant le modèle. On a ainsi une approche dans le domaine JPEG équivalente à celle au niveau des photosites quant à la sécurité empirique. On peut également observer (tableau 1) l'impact de l'hypothèse d'indépendance sur la réalisation du bruit sur les coefficients DCT.

JPEG QF	Pseudo insertion	4-treillis	Insertion indépendante	SI-Uniward 1 bpnzac
95	40.9	41.2	0.8	0.4
85	41.9	41.2	10.8	12.3
75	41.3	41.6	27.0	24.8

TABLE 1 – $P_E(\%)$ pour différents QF et stratégies d'insertion sur E1Base.

4.2 Influence de l'alphabet d'insertion (dynamique du signal stego) :

L'impact de la taille de l'alphabet $[-K, ..., 0, ...K]$ est ici mesuré au regard de la sécurité empirique. Le tableau présente les résultats obtenus en fonction de QF.

Dans le cas d'une insertion ternaire ($[-1, 0, +1]$) le schéma devient très détectable à QF $= 95$. Afin d'atteindre une sécurité comparable à l'emploi d'un alphabet sans limite de taille il faut utiliser au moins 7 symboles : $[-3, ..0, .., +3]$. Au contraire, à cause de la quantification une insertion ternaire convient pour QF $= 75$.

JPEG QF	$K = 1$	$K = 2$	$K = 3$
100	1.0	12.9	28.7
95	3.5	23.6	39.3
85	39.8	39.8	39.8
75	40.4	40.4	40.4

TABLE 2 – $P_E(\%)$ pour différents QF et en fonction de la taille de l'alphabet K pour une pseudo-insertion.

5 Conclusions et perspectives :

Ce papier s'inscrit comme une synthèse de nos avancées récentes [8, 2] dans le domaine de la stéganographie naturelle. Nous avons montré la faisabilité d'une insertion ne perturbant pas les dépendances entre les coefficients DCT pour un développement particulier. Il est possible qu'une approche similaire puisse s'appliquer à d'autres développements linéaires : tel sera l'objet de nos travaux futurs en addition d'une stéganalyse contre des réseaux de neurones.

Références

[1] P. Bas, "Steganography via Cover-Source Switching," 2016, IEEE Workshop on Information Forensics and Security (WIFS).

[2] T. Denemark, P. Bas, and J. Fridrich, "Natural Steganography in JPEG Compressed Images," in *Electronic Imaging*, San Francisco, United States, 2018.

[3] T. Taburet, P. Bas, W. Sawaya, and J. Fridrich, "Computing Dependencies between DCT Coefficients for Natural Steganography in JPEG Domain," in *7th ACM Workshop on Information Hiding and Multimedia Security*, 2019.

[4] W. Li, W. Zhang, K. Chen, W. Zhou, and N. Yu, "Defining joint distortion for jpeg steganography," in *Proceedings of the 6th ACM Workshop on Information Hiding and Multimedia Security*. ACM, 2018, pp. 5–16.

[5] V. Holub and J. Fridrich, "Low-complexity features for jpeg steganalysis using undecimated dct," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 2, pp. 219–228, 2015.

[6] R. Cogranne, V. Sedighi, J. Fridrich, and T. Pevnỳ, "Is ensemble classifier needed for steganalysis in high-dimensional feature spaces ?" in *Information Forensics and Security (WIFS), 2015 IEEE International Workshop on*. IEEE, 2015, pp. 1–6.

[7] V. Holub, J. Fridrich, and T. Denemark, "Universal distortion function for steganography in an arbitrary domain," *EURASIP Journal on Information Security*, vol. 2014, no. 1, pp. 1–13, 2014.

[8] T. Taburet, P. Bas, W. Sawaya, and J. Fridrich, "A Natural Steganography Embedding Scheme Dedicated to Color Sensors in the JPEG Domain," in *Electronic Imaging*, Burlingame, United States, Jan. 2019.