Utilisation du bruit photonique pour la stéganographie

Patrick BAS

Univ. Lille, CNRS, Centrale Lille, UMR 9189
CRIStAL , F-59000 Lille, France
Patrick.Bas@ec-lille.fr

Résumé – Dans ce papier nous proposons d'utiliser, via un algorithme appelé " Stéganographie Naturelle" (NS), la modélisation du changement de sensibilité ISO afin d'insérer de manière indétectable un message de taille importante dans une image numérique. En première approximation ce changement de sensibilité peut être modélisé par l'ajout d'un bruit photonique indépendant de loie normale sur les photosites du capteur. L'implémentation du schéma NS cherche à modéliser la distribution de ce signal stéganographique après développement d'une image RAW vers une image pouvant être éditée. Les étapes de développement considérées sont la quantification sur 8 bits, la correction gamma et le redimensionnement d'un facteur 2. Même si ces premières implémentations sont loin de prendre en compte tous les développements possibles, dans chacun des cas, nous montrons que cette modélisation statistique permet de garantir une indétectabilité bien supérieure à celle des méthodes classiques.

Abstract – This paper proposes to model the ISO switch occurring in digital images acquisition to embed an undetectable high payload by a scheme called "Natural Steganography" (NS). This sensitivity change can be modeled by addition of a photonique noise independently and normally distributed on each sensor photo-site, and the implementation of a NS scheme looks at modeling the steganographic signal at the very end of the development pipeline. We consider here 8 bits quantization, gamma correction and downscaling by a factor 2. Even if we are far from taking into account all the possible development pipelines, we show that these statistical models enable to guaranty an undetectability more important than traditional schemes.

1 Introduction

La stéganographie cherche à insérer de manière indétectable un message de taille la plus importante possible dans un contenu anodin. Nous nous focaliserons ici sur les images numériques, car ce contenu est majoritairement utilisé lors des échanges numériques. L'indétectabilité est pratiquement garantie par l'impossibilité pour le stéganalyste (la personne effectuant l'analyse stéganographique, aussi appelée stéganalyse) de distinguer une image appelée *cover* (une images de couverture, ne contenant pas d'information cachée) d'une image *stégo* (une image modifiée par un algorithme de stéganographie).

D'un point de vue formel, si chaque élément de l'objet *cover* $x_{i,j}$ (ici nous considérerons des pixels, mais cela sera par exemple des coefficients DCT pour la stéganographie destinée aux images JPEG) est modifié en ajoutant la valeur $k \in \mathbb{N}$ pour produire l'échantillon *stégo* $y_{i,j} = x_{i,j} + k$, et en appelant $\pi_{i,j}(k)$ la probabilité d'ajouter k sur l'échantillon $x_{i,j}$, la taille du message inséré sera égale à l'entropie du signal stéganographique, soit[1] :

$$H = -\sum_{i,j,k} \pi_{i,j}(k) \log_2 \pi_{i,j}(k) \,\text{bits}, \tag{1}$$

en supposant qu'un codage optimal existe et que la loi $\pi_{i,j}(k)$ soit indépendamment distribuée pour chaque pixel. D'un point de vue pratique, un codage quasi-optimal peut être obtenu via l'utilisation de STC (syndrom trellis codes) [1].

Un schéma de stéganographie cherche donc à trouver la loi $\pi_{i,j}(k)$ permettant, pour un message de taille donné, de minimiser la détectabité de l'image *stégo*.

Sans rentrer dans les détails par faute de place, la majorité des schémas actuels de stéganographie ne considèrent qu'une insertion binaire ou ternaire ($k \in \{-1, 0, 1\}$) et les probabilités $\pi_{i,j}(k)$ sont calculées à partir de coûts heuristiques traduisant une détectabilité propre à chaque échantillon. Un pixel se trouvant par exemple dans une zone homogène de l'image sera considéré comme plus "détectable" qu'un pixel se trouvant dans une zone texturée et verra sont coût plus important. Ce principe est la base de schémas tels que la famille de schémas Uniward [2] ou encore MiPOD [3], même si ce dernier repose sur un modèle statistique plus fin. D'autres astuces permettent d'améliorer la sécurité d'un schéma de stéganographie, à savoir l'utilisation d'une image "pre-*cover*", c'est à dire avant la quantification finale générant l'image *cover* [4], la synchronisation partielle des modifications dues à l'insertion [5, 6] ou encore des mécanismes d'insertions itératifs venant séquentiellement défier le stéganalyste [7].

2 Insertion par changement de source

L'idée ici n'est plus de prendre en compte le contenu de l'image afin de garantir l'indétectabilité, mais d'utiliser le bruit d'acquisition photonique et le principe de *changement de source cover*. Par *source* nous entendons ici la distribution statistique d'images provenant d'un capteur d'Appareil Photos Numérique

(APN) donné (ou d'un modèle de capteur) à une sensibilité ISO donnée. Le principe du "changement de source cover", initialement présenté dans [8], propose d'effectuer l'insertion du message en mimant une source cover différente de la source initiale, c'est à dire une sensibilité ISO différente. Le signal permettant de passer d'un modèle de source $\mathcal{S}_1$ à un autre modèle $\mathcal{S}_2$ sera le signal stéganographique. Puisque reposant sur les propriétés naturelles du bruit photonique, nous avons choisi d'appeler ce schéma la "Stéganographie Naturelle" (ou NS pour "Natural Steganography").

Plus formellement, pour un paramètre ISO ISO_1 donné, le bruit photonique $N_{i,j}^{(1)}$ appliqué sur chaque photosite de position (i,j) peut être approximé par un bruit indépendant distribué selon une loi normale et dont la variance est directement en relation affine avec la valeur moyenne du photo-site $\mu_{i,j}$:

$$N_{i,j}^{(1)} \sim \mathcal{N}(0, a_1\mu_{i,j} + b_1). \tag{2}$$

Ainsi, chaque photosite $x_{i,j}^{(1)}$ peut être modélisé par $x_{i,j}^{(1)} = \mu_{i,j} + n_{i,j}^{(1)}$ et est donc distribué selon $X^{(1)} \sim \mathcal{N}(\mu_{i,j}, a_1\mu_{i,j} + b_1)$.

L'insertion NS consiste donc a générer le signal stéganographique de manière à mimer le passage de la sensibilité ISO_1 à une sensibilité ISO_2 (puisque nous ajoutons un signal $ISO_2 > ISO_1$). Nous avons donc ici $N_{i,j}^{(2)} \sim \mathcal{N}(0, a_2\mu_{i,j} + b_2)$ et $x_{i,j}^{(2)} = \mu_{i,j} + n_{i,j}^{(2)}$. D'un point de vue pratique, puisque l'ajout de deux variables normales indépendantes reste une variable normale indépendante, et en supposant que $\mu_{i,j} \simeq x_{i,j}^{(1)}$, le signal stéganographie $S_{i,j}$ sera alors distribué selon :

$$S_{i,j} \sim \mathcal{N}(0, (a_2 - a_1)x_{i,j}^{(1)} + b_2 - b_1). \tag{3}$$

En utilisant les notations suivantes : $a' \triangleq a_2 - a_1$, $b' \triangleq b_2 - b_1$, $\sigma_S^2 \triangleq a'x_{i,j}^{(1)} + b'$, le photosite de l'image *stégo* $y_{i,j}$ est alors distribué comme :

$$Y_{i,j} \sim \mathcal{N}(x_{i,j}^{(1)}, \sigma_S^2). \tag{4}$$

Notons que cette modélisation s'effectue sur des v.a. continues et que l'insertion réelle doit s'effectuer après quantification, c.a.d. sur des variables aléatoires discrètes comme expliqué dans la section suivante.

3 Insertion et prise en compte de divers traitements

Le principe d'insertion présenté dans la section précédente ne s'effectue qu'au niveau du photosite codé par l'image RAW, nous étudions ici comment répercuter ces différents traitements sur l'image développée, c'est à dire destinée à être éditée.

3.1 Aucun post-traitement

Si aucun post-traitement n'est effectué, la seule opération qui convient de prendre en compte est la quantification des pixels sur 8 bits. Nous cherchons donc à calculer les probabilités de modification $\pi_{i,j}(k) = \Pr[\bar{S}_{i,j} = k]$ lié à la quantification, où la v.a. $\bar{S}$ représente le pixel quantifié. En partant de la valeur codée sur 16 bits du photosite x_{16B}, et de la valeur quantifiée sur $x_{8B} = Q_\Delta(x_{16B})$ où $Q_\Delta(.)$ représente la fonction de quantification, nous obtenons :

$$\pi(k) \quad = \quad \frac{1}{2}\left(\operatorname{erf}\left(\frac{u_{k+1}-x_{16B}}{\sqrt{2\sigma_S^2}}\right) - \operatorname{erf}\left(\frac{u_k-x_{16B}}{\sqrt{2\sigma_S^2}}\right)\right), \tag{5}$$

avec $u_k = x_{8B} - (0.5 - k)\Delta$. La taille du message inséré se calcule ensuite à partir de l'entropie définie en (1). Le calcul des $\pi(k)$ pour chaque pixel permet ensuite d'utiliser un codeur STC qui insérera le message désiré en utilisant des poids directement calculés à partir des $\pi(k)$ (voir [1]).

3.2 Correction gamma

Puisque la correction gamma est une opération modifiant chaque photosite de façon indépendante, un schéma NS peut facilement prendre en compte cette opération, en remplaçant la distribution normale du signal stéganographique par la distribution après correction Gamma.

Pour des raisons de manque de place, nous renvoyons le lecteur à [8] qui donne l'approximation possible lorsque signal stéganographique est de puissance faible, ce qui est souvent le cas.

3.3 Redimensionnement

La prise en compte du redimensionnement est plus complexe et est décrite en détails pour différents types de redimensionnement dans [9]. Lorsqu'il s'agit par exemple d'un redimensionnement bilinéaire, la procédure d'insertion ne se contente plus d'insérer le message en considérant les v.a. $S_{i,j}$ comme des variables indépendantes. Comme le montre la Figure 1-(a), l'étape de convolution suivie du sous-échantillonnage entraîne une dépendance entre deux pixels développés voisins, alors que deux pixels non voisins pourront avoir un signal stéganographique tiré indépendamment pour chaque pixel.

La procédure d'insertion repose sur l'utilisation de quatre treillis disjoints illustrés sur la Figure 1-(b). Une portion du message est d'abord insérée sur le premier treillis $\mathcal{E}_1$, puis de manière séquentielle sur les treillis $\mathcal{E}_2$, $\mathcal{E}_3$ et $\mathcal{E}_4$. L'insertion se décompose donc en 4 étapes :

- **Étape 1 :** Sur le treillis $\mathcal{E}_1$, les pixels développés sont modifiés par un signal stéganographique $\bar{S}_{\mathcal{E}_1} \sim \mathcal{N}(0, \sigma_{S1}^2)$ avec $\sigma_{S1}^2 = \sum_{i,j} c_{i,j}^2 \sigma_S^2(i,j)$, $c_{i,j}$ représentant les coefficients du filtre bi-linéaire. Les valeurs du bruit du capteur sur chaque photosite (les positions grisées identiquement sur la Figure 1-(b)) sont également tirées conditionnellement au tirage de $S_{\mathcal{E}_1}$ et des valeurs précédemment tirées.

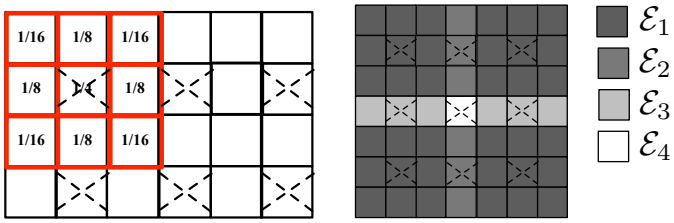
- **Étape 2 :** Ces deux opérations (tirage du signal stéganographique dans le domaine développé puis tirage du bruit re-

latif aux photosites) sont répétées sur le treillis $\mathcal{E}_2$ mais cette fois-ci en prenant en compte les valeurs des photosites tirés sur $\mathcal{E}_1$ pour tirer $\bar{S}_{\mathcal{E}_2}$.

- **Étapes 3 et 4 :** Nous répétons l'opération précédente sur $\mathcal{E}_3$ et sur $\mathcal{E}_4$ en prenant en compte les valeurs des photosites tirés précédemment.

Comme précédemment l'insertion réelle du message s'effectue en calculant les probabilités de modification après quantification et ce pour chaque treillis, et en utilisant un codage STC multi-couche, dont les poids sont calculés à partir des probabilités de modification.

Nous pouvons remarquer que les étapes 2, 3 et 4 permettent naturellement de synchroniser (c'est à dire de corréler) les changements effectués entre le premier treillis et les autres. Nous retrouvons ici un mécanisme utilisé de manière heuristique dans [5, 6] pour améliorer l'indétectabilité d'un schéma de stéganographie.



(a) (b)

FIGURE 1 – (a) : Noyau utilisé pour le redimensionnement bilinéaire d'un facteur 2 et dépendance des pixels développés (représentés par des croix). (b) : les quatre treillis $\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3, \mathcal{E}_4$ utilisés lors de l'insertion, colorés en fonctions des photosites qui sont échantillonnés lors de l'insertion.

4 Résultats

Afin d'évaluer le principe de l'insertion par changement de source, nous avons eu besoin d'acquérir des bases d'images à deux sensibilités ISO différentes. Afin de s'affranchir de l'étape de dématriçage lors du développement nous avons utilisé un Leica M Monochrome Type 230, pour ensuite prendre 172 photos à chaque sensibilité (ici ISO 1000 et ISO 1250). Chaque photo a été ensuite re-découpée en images disjointes de taille 512×512. Cette base d'images, appelée MonoBase, peut être téléchargée [10] et le code source présentant le système d'insertion est téléchargeable sur Github [11].

4.1 Procédures d'évaluation

Nous utilisons les caractéristiques spatiales (SRM) [12] associé à un classifieur par ensemble [13] et nous adoptons classiquement comme métrique l'erreur moyenne totale

$$P_E = \min((P_{FA} + P_{MD})/2),$$

moyennée sur 10 découpages où les bases d'apprentissage et de test sont divisées en moitiés égales.

Contrairement à une évaluation classique en stéganalyse qui cherche à différencier les images *cover* des images *stégo* directement générées à partir des *cover*, nous cherchons ici à évaluer la capacité du schéma NS à changer de source et nous confrontons donc des images *cover* acquises à ISO 1250 à des images *stégo* générées à partir d'images *cover* acquises à ISO 1000 (mais mimant une sensibilité ISO 1250).

Nous avons utilisé les valeurs $a' = 2.1\,10^{-5}$ et $b' = 8.4\,10^{-7}$ sur des valeurs de photosites normalisées à 1. La procédure permettant d'estimer ces paramètres est détaillée dans [8], elle repose sur plusieurs acquisitions d'une même scène comportant une transition d'un noir au blanc. Le taux d'insertion est directement calculé via la formule (1) après calcul des $\pi(k)$.

À titre de comparaison, nous comparons notre schéma au schéma S-Uniward utilisé ici à ISO 1000 dans sa version de base et dans sa version utilisant l'image pre-*cover* codée sur 16 bits (S-Uniward-SI)[2].

4.2 Détectabilité après les différents traitements

Nous évaluons tour à tour chacun des traitements.

Le tableau 1 compare la détectabilité du schéma NS à celle des schémas S-Uniward et S-Uniward-SI pour un taux d'insertion moyen de $1.24\,\text{bpp}$ pour NS et un taux constant de $1.24\,\text{bpp}$ pour les deux autres schémas. Nous pouvons remarquer que même dans sa version utilisant la pré-*cover*, la famille S-Uniward est bien plus détectable que notre implémentation.

	NS	SUni-SI	SUni
P_E	**44.8%**	18.2%	12.3%

TABLE 1 – Comparaison avec S-Uniward et S-Uniward-SI à $1.24\,\text{bpp}$.

Le tableau 2 montre que la correction gamma lorsqu'elle est prise en compte lors de la génération du signal stéganographique, n'a aucun impact sur la détectabilité de la méthode NS.

γ	2.5	2	1.5	1	0.5
P_E	44.4%	44.5%	43.7%	44.8%	46.2%
E_r(bpp)	1.61	1.62	1.55	1.24	0.5

TABLE 2 – Détectabilité après correction Gamma.

Le tableau 3 évalue la détectabilité de notre méthode d'une part lorsque le signal stéganographique est généré via l'algorithme décrit en section 3.3, d'autre part lorsque le signal stéganographique est tiré indépendamment selon la loi affectée à $\bar{S}_{\mathcal{E}_1}$, et permet d'illustrer la nécessité de générer un signal stéganographique ayant un modèle statistique proche de celui de la source.

Enfin, la figure 2 présente l'évolution du taux d'insertion en fonction du facteur d'échelle lors du redimensionnement et ce pour différents noyaux d'interpolations (sous échantillon-

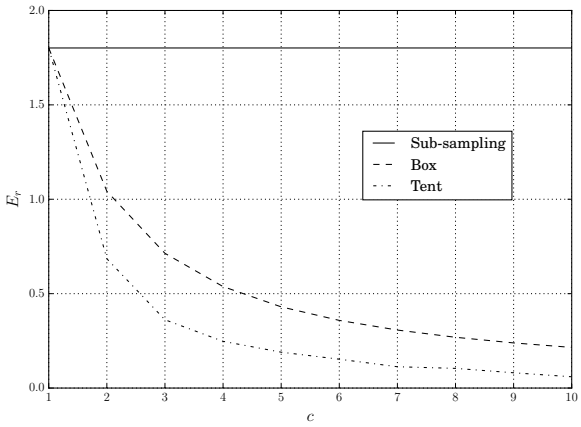


FIGURE 2 – Évolution du taux d'insertion vs facteur d'échelle.

nage naïf, noyau *box* (moyennage) et noyau *tent* (bilinéaire). Nous pouvons constater d'une part la logique décroissance du taux d'insertion en fonction du facteur d'échelle (hormis pour le sous-échantillonnage), d'autre part le taux d'insertion plus faible du noyau bilinéaire par rapport aux deux autres. Notons que ces résultats sont en accord avec les résultats pratiques présentés sur d'autres schémas d'insertion dans [14].

	NS	NS (Bilin.)	NS (Indép.)
P_E	44.8%	48.0%	22.6%

TABLE 3 – Détectabilité après re-dimensionnement d'un facteur 2.

5 Conclusions et perspectives

Nous avons présenté dans ce papier une nouvelle méthode de stéganographie qui utilise le bruit photonique et le concept de changement de source afin d'insérer un message de taille importante tout en garantissant l'indétectabilité après insertion. Nous avons montré également que ce schéma pouvait supporter différents types de développements tels que la quantification vers 8 bits, la correction gamma, ou encore le redimensionnement de l'images. Des travaux sont en cours afin de prendre en compte la transformation DCT et ainsi d'assurer une insertion sur des images JPEG.

Nos travaux futurs vont chercher à prendre en compte toute la chaîne de développement d'une image classique, c'est à dire le dé-matriçage, les transformations couleurs, et possiblement des transformations non-linéaires.

Enfin, dans une démarche adversarielle, nous chercherons également à construire des caractéristiques plus sensibles à notre signal stéganographique. Notons que le problème présenté ici est proche d'un problème de *forensics*, et que s'il est chose aisée de modifier les données EXIF renseignant la sensibilité d'une image, il existe peut être des méthodes d'analyse capables d'extraire une empreinte liée à la sensibilité ISO au sein du signal.

Références

[1] T. Filler, J. Judas, and J. Fridrich, "Minimizing additive distortion in steganography using syndrome-trellis codes," *Information Forensics and Security, IEEE Transactions on*, vol. 6, no. 3, pp. 920–935, 2011.

[2] 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.

[3] V. Sedighi, R. Cogranne, and J. Fridrich, "Content-adaptive steganography by minimizing statistical detectability," *Information Forensics and Security, IEEE Transactions on*, vol. 11, no. 2, pp. 221–234, 2016.

[4] T. Denemark and J. Fridrich, "Side-informed steganography with additive distortion," in *Information Forensics and Security (WIFS), 2015 IEEE International Workshop on*. IEEE, 2015, pp. 1–6.

[5] ——, "Improving steganographic security by synchronizing the selection channel," in *Proceedings of the 3rd ACM Workshop on Information Hiding and Multimedia Security*. ACM, 2015, pp. 5–14.

[6] B. Li, M. Wang, X. Li, S. Tan, and J. Huang, "A strategy of clustering modification directions in spatial image steganography," *Information Forensics and Security, IEEE Transactions on*, vol. 10, no. 9, pp. 1905–1917, 2015.

[7] S. Kouider, M. Chaumont, and W. Puech, "Adaptive steganography by oracle (aso)," in *Multimedia and Expo (ICME), 2013 IEEE International Conference on*. IEEE, 2013, pp. 1–6.

[8] P. Bas, "Steganography via Cover-Source Switching," 2016, IEEE Workshop on Information Forensics and Security (WIFS). [Online]. Available : https ://hal.archives-ouvertes.fr/hal-01360024

[9] ——, "An embedding mechanism for Natural Steganography after down-sampling," 2017, IEEE ICASSP.

[10] ——, "Monobase," http ://patrickbas.ec-lille.fr/MonoBase/, July 2016.

[11] ——, "Source Codes for Natural Steganography," https ://github.com/patrickbas/NaturalSteganography, October 2016.

[12] J. Fridrich and J. Kodovsky, "Rich models for steganalysis of digital images," *Information Forensics and Security, IEEE Transactions on*, vol. 7, no. 3, pp. 868–882, 2012.

[13] J. Kodovsky, J. Fridrich, and V. Holub, "Ensemble classifiers for steganalysis of digital media," *Information Forensics and Security, IEEE Transactions on*, vol. 7, no. 2, pp. 432–444, 2012.

[14] J. Kodovsky and J. Fridrich, "Steganalysis in resized images," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. IEEE, 2013, pp. 2857–2861.