



HAL
open science

Détection Optimale d'Information Cachées Indépendante du Contenu des Images

Rémi Cogranne, Lionel Fillatre, Florent Retraint, Cathel Zitzmann

► **To cite this version:**

Rémi Cogranne, Lionel Fillatre, Florent Retraint, Cathel Zitzmann. Détection Optimale d'Information Cachées Indépendante du Contenu des Images. Journées codes et stéganographie, Jan 2011, Saint Cyr - Coëtquidan, France. pp.6. hal-00640667

HAL Id: hal-00640667

<https://hal.science/hal-00640667v1>

Submitted on 14 Nov 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Détection Optimale d'Information Cachées Indépendante du Contenu des Images

Rémi Cogranne, Lionel Fillatre, Florent Retraint et Cathel Zitzmann

Laboratoire LM2S/STMR, UMR CNRS 6279, Université de technologie de Troyes.

Abstract. Dans ce papier, une nouvelle méthodologie est proposée pour la détection de stéganographie par substitution des bits de poids faibles (LSB) d'une image naturelle. Une modélisation du système imageur et de la scène permet la construction d'un modèle linéaire localement adapté au contenu des images. La détection est formalisée comme un test statistique d'hypothèses. Deux tests sont proposés, le test optimal théorique du rapport de vraisemblance (RV) et un test pratique quasi-optimal. Les apports de cette méthodologie sont 1) la prise de décision est indépendante du contenu de l'image, 2) les performances du test optimal sont explicitées et peuvent servir de "benchmark" et 3) les performances du test quasi-optimal sont bornées, ainsi la perte de puissance est quantifiée. Les résultats présentés confirment la pertinence des recherches théoriques.

Keywords: Modèle local d'image, approximation linéaire par morceaux, stéganalyse, test statistique d'hypothèses.

1 Introduction

Dans le domaine de la stéganalyse, la majorité des méthodes actuelles de détection reposent sur l'apprentissage supervisés [2,3]. De plus en plus de descripteurs sont utilisés pour améliorer la stéganalyse par : 1) une diminution d'une erreur de classification (souvent la somme des fausses-alarmes et des non-détection), 2) la détection de tous les schémas de stéganographie (stéganalyse dite universelle) et 3) la détection dans tous les types d'images, naturelles ou infographiques.

La faiblesse de cette approche, outre le choix des descripteurs, réside dans la manque de modèles statistiques des descripteurs. les inconvénient qui en découlent sont : 1) les résultats obtenus dépendent des données d'apprentissage et de test et 2) aucune garantie ne peut être donnée sur les performances (en termes de fausses-alarmes notamment).

La plupart des descripteurs utilisés pour l'apprentissage supervisé se basent implicitement sur des modèles d'images (champs de Markov, ondelettes, ...) mais peu de travaux dans la littérature exploitent explicitement ces modèles. En effet la difficulté de cette approche est liée au fait qu'il faut disposer 1) d'un modèle statistique précis des images 2) d'un modèle du contenu des images. Ces deux modèles sont nécessaires, le premier permet d'établir le test d'hypothèses et le second permet au test d'être indépendant du contenu de l'image.

Cette dualité est la motivation qui est au cœur de cette article : créer une passerelle entre la modélisation des images et la stéganalyse. La méthodologie proposée repose d'une part sur un modèle localement adapté des images naturelles et d'autre part sur la théorie de la décision statistique. Les apports sont, en principes, les suivants : 1) la décision est prise indépendamment du contenu de l'image, considéré comme un "paramètre de nuisance" ; 2) le test peut satisfaire des contraintes sur le taux de fausses-alarms en fixant un seuil de détection indépendant des données. 3) les performances du test optimal du rapport de vraisemblance (RV) sont donnée asymptotiquement. La perte d'optimalité d'un test pratique peut être quantifiée.

2 Modélisation d'une image naturelle

Dans cette section, un modèle de la chaîne d'acquisition puis de la scène sont présentés pour aboutir à un modèle local du contenu des images naturelles.

2.1 Processus d'acquisition d'une image

Une scène est décrite par la luminance spectrale incidente sur le système optique $B(x, y, \lambda)$ où $(x, y) \in [0, x_M] \times [0, y_M] = \mathcal{X} \subset \mathbb{R}^2$ représentent les coordonnées continues de la scène mise à l'échelle du capteur. Pour un appareil photo, cette grandeur est mesurée pour les trois canaux de couleurs $c = \{r, v, b\}$ par :

$$B_c(x, y) = \int_{R^+} T_c(\lambda) B(x, y, \lambda) d\lambda$$

où $T_c(\lambda)$ représente les filtres couleurs, l'efficacité quantique du capteur et l'énergie des photons. Les canaux étant traités indifféremment, l'indice c est omit.

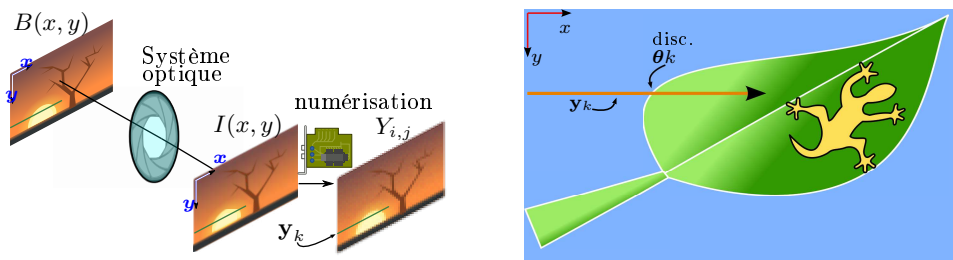
Dans des conditions d'acquisitions stationnaires, le canal optique se modélise par un noyau de convolution $h(x, y)$ [7]. Ce noyau dépend des conditions d'acquisition (système optique, atmosphère, etc...) et de nombreuses modélisations sont proposés dans la littérature. L'irradiance du capteur, nombre de photo-électrons générés par unité de surface, est donné par [7, 8] :

$$I(x, y) = (h * B)(x, y) \stackrel{\text{def.}}{=} \iint_{\mathbb{R}^2} B(x-u, y-v) h(u, v) du dv \quad (1)$$

avec $*$ l'opérateur de convolution-2D. Ce papier considère uniquement le cas du noyau Gaussien isotropique. Enfin, l'irradiance $I(x, y)$ est intégrée sur la surface des pixels $[0, s_x] \times [0, s_y]$ durant un temps τ [8] :

$$\omega_{i,j} = \tau \cdot \int_{x_i}^{x_i+s_x} \int_{y_i}^{y_i+s_y} I(u, v) S_r(u, v) dv du \quad (2)$$

où il est supposé dans un premier temps que $S_r(x, y)$, la réponse et l'amplification du capteur, est constante et que les pixels sont contigus $x_{i+1} - x_i = s_x$ et $y_{j+1} - y_j = s_y$.



(a) Une chaîne d'acquisition d'image naturelles. (b) Une scène en 2D et sa modélisation 1D.

Fig. 1: Les descriptions de la scène et de l'acquisition dans le modèle d'image.

2.2 Modélisation de l'image numérisée sous forme de segment

Pour modéliser la scène imagée, l'approche de [11] est utilisée : la scène est assimilée à une juxtaposition d'objets dont la radiance est continue. Les frontières entre ces objets introduisent des discontinuités dans $B(x, y)$. Ainsi, la scène admet la décomposition : $B(x, y) = B_c(x, y) + B_s(x, y)$, où $B_c(x, y)$ est continue et $B_s(x, y)$, correspondant aux frontières entre les objets, est discontinues.

Le modèle précédent n'est pas exploitable en pratique car les frontières ne sont pas connues et ne peuvent en aucun cas être modélisées de façon paramétrique. Or la détection d'informations cachées nécessite de disposer d'un modèle local adapté aux images naturelles qui soit simple. Dans ce but, l'image numérique est découpée en K segments de N pixels contigus de la même ligne (ou colonne).

En notant $\mathbf{z}_k = (z_{k,1}, \dots, z_{k,M})^T$ les vecteur de M composantes, le modèle de la scène permet l'expression de (2) par segment sous la forme algébrique:

$$\boldsymbol{\omega}_k = \mathbf{H}_k \mathbf{s}_k + \mathbf{F}(\boldsymbol{\theta}_k) \mathbf{u}_k. \quad (3)$$

La matrice \mathbf{H}_k de taille $N \times p$ correspond à la partie continue du k -ième segment modéliser par un polynôme de degré $p - 1$, dont le vecteur \mathbf{s}_k contient les coefficients. La matrice $\mathbf{F}(\boldsymbol{\theta}_k)$ de taille $N \times r_k$ modélise les r_k discontinuités, paramétrées par les vecteurs $\boldsymbol{\theta}_k = (\theta_{k,1}, \dots, \theta_{k,r_k})^T$ et \mathbf{u}_k , dans le k -ième segment. Dans la suite, sans perte de généralité, la base polynomiale est choisie identique sur tous les segments : $\mathbf{H}_k = \mathbf{H}$ et ainsi $\boldsymbol{\omega}_k \in \mathbf{v}_k = \text{im}(\mathbf{H}) \cup \text{im}(\mathbf{F}(\boldsymbol{\theta}_k)) \subset \mathbb{R}^N$.

Finalement, $\mathbf{y}_k \in \mathcal{Y}^N$, le k -ième segment, peut s'écrire [5] : $\mathbf{y}_k = \mathcal{Q}_\Delta(\boldsymbol{\omega}_k + \boldsymbol{\xi}_k)$, où $\boldsymbol{\xi}_k \sim \mathcal{N}(0; \Sigma_k)$, $\Sigma_k = \text{diag}(\sigma_{k,1}^2, \dots, \sigma_{k,N}^2)$ et $\mathcal{Q}_\Delta : \mathbb{R} \mapsto \mathcal{Y} = \{0, \dots, \ell-1\}$ est le quantificateur scalaire uniforme de pas Δ appliqué sur chaque pixel.

3 Formalisation du problème de la stéganalyse

A partir de la section 2.2, la distribution de probabilité des pixels d'une image naturelle est donné par $\mathcal{P}_{\omega_{k,n}} = \{P_{\omega_{k,n}}(0), \dots, P_{\omega_{k,n}}(\ell-1)\}$ où $\forall m \in \mathcal{Y}$:

$$\mathcal{P}_{\omega_{k,n}}(m) = \mathbb{P}[y_{k,n} = m | \omega_{k,n}, \sigma_{k,n}] = \frac{1}{\sigma_{k,n}} \int_{\Delta(m-\frac{1}{2})}^{\Delta(m+\frac{1}{2})} \phi\left(\frac{x - \omega_{k,n}}{\sigma_{k,n}}\right) dx \quad (4)$$

Pour modéliser statistiquement la stéganographie dans les LSB, on définit le taux d'insertion comme le nombre de bits insérés par pixel de l'image. Il est souvent admis que les bits du message, 0 ou 1, sont équiprobables [4, 6]. De la même façon, la distribution de $y_{k,n}$ après insertion d'un message par substitution de LSB est alors donnée par $\mathcal{Q}_{\omega_{k,n}}^R = \{Q_{\omega_{k,n}}^R(0), \dots, Q_{\omega_{k,n}}^R(\ell-1)\}$ où $Q_{\omega_{k,n}}^R(m)$ vaut [4]:

$$Q_{\omega_{k,n}}^R(m) = \left(1 - \frac{R}{2}\right) P_{\omega_{k,n}}(m) + \frac{R}{2} P_{\omega_{k,n}}(m - \eta(m)) \quad (5)$$

avec $\eta(m) = -1$ si m est pair et $\eta(m) = 1$ si m est impair.

La stéganalyse peut se formaliser comme un test statistique entre deux hypothèses composite visant à décider si :

$$\begin{cases} \mathcal{H}_0 = \{\mathbf{y}_{k,n} \sim \mathcal{P}_{\omega_{k,n}}, \forall k = 1 \dots, K, \forall n = 1 \dots, N\} \\ \mathcal{H}_1 = \{\mathbf{y}_{k,n} \sim \mathcal{Q}_{\omega_{k,n}}^R, \forall k = 1 \dots, K, \forall n = 1 \dots, N, \forall R_0 \leq R \leq 1\} \end{cases}$$

où R_0 un taux minimal d'insertion (raisonnable). Le but est de trouver un test $\varphi: \mathcal{Y}^{K \cdot N} \mapsto \{\mathcal{H}_0; \mathcal{H}_1\}$ acceptant l'hypothèse \mathcal{H}_i si $\varphi(\mathbf{Y}) = \mathcal{H}_i$ (cf. [10] pour plus d'informations sur les tests statistiques). Soit

$$\mathcal{K}_\alpha = \left\{ \varphi : \sup_{\boldsymbol{\Omega}_K \in \boldsymbol{\Upsilon}_K} \mathbb{P}[\varphi(\mathbf{Y}) = \mathcal{H}_1 | \mathcal{H}_0] \leq \alpha \right\}$$

la classe des tests dont la probabilité de fausses alarmes est majorée par α , où $\boldsymbol{\Omega}_K = (\omega_1, \dots, \omega_K)$ appartient au sous-espace $\boldsymbol{\Upsilon}_K = \mathbf{v}_1 \times \dots \times \mathbf{v}_K$. La fonction de puissance $\beta_\varphi(\boldsymbol{\Omega}_K, R)$, est la probabilité de détection du message caché $\beta_\varphi(\boldsymbol{\Omega}_K, R) = \mathbb{P}(\varphi(\mathbf{Y}) = \mathcal{H}_1 | \mathcal{H}_1)$. Le but est alors de maximiser $\beta_\varphi(\boldsymbol{\Omega}_K, R)$ uniformément par rapport R en tenant compte de $\boldsymbol{\Omega}_K \in \boldsymbol{\Upsilon}_K$ qui est un paramètre de nuisance pour la détection.

4 Tests optimaux pour la stéganalyse

Par manque de place, les détails sur les tests et les expérimentations seront données dans la présentation orale. Cette section présente les résultats de détection obtenus par deux tests statistiques originaux. Lorsque $\boldsymbol{\Omega}_K$ est connu, le test optimal du RV, ou test le plus puissant dans la classe \mathcal{K}_α , est donné par :

$$\delta_{NP}(\mathbf{Y}) = \begin{cases} \mathcal{H}_0 & \text{si } \Lambda_{NP}(\mathbf{Y}) < \tau_\alpha \\ \mathcal{H}_1 & \text{si } \Lambda_{NP}(\mathbf{Y}) \geq \tau_\alpha \end{cases} \quad \text{où } \Lambda_{NP}(\mathbf{Y}) = \sum_{k=1}^K \sum_{n=1}^N \ln \left(\frac{\mathcal{Q}_{\omega_{k,n}}^R(y_{k,n})}{\mathcal{P}_{\omega_{k,n}}(y_{k,n})} \right) \quad (6)$$

où τ_α est choisi de sorte que $\delta_{NP} \in \mathcal{K}_\alpha$. Il peut être montré que l'expression de $\Lambda_{NP}(\mathbf{Y})$ ne dépend pas de R ; δ_{NP} est donc optimal pour tout $R > 0$. La puissance de ce test peut être calculée asymptotiquement de façon précise.

Lorsque $\boldsymbol{\Omega}_K$ est inconnu, il est proposé d'utiliser le test $\hat{\delta}$ du RV généralisé (RVG) ; ce dernier utilise le test du RV (6) en remplaçant $\boldsymbol{\Omega}_K$ par une estimation

$\widehat{\Omega}_K$. Cette estimation $\widehat{\Omega}_K$ est possible à partir du modèle proposé dans (3). La perte d'optimalité est quantifiée par le calcul de bornes de puissance de ce test.

Les calculs de performances de ces tests sont en adéquation avec la loi de la racine carrée de la stéganographie [9]. Ces calculs ne pouvant être détaillés, ils seront présentés oralement. Les deux tests proposés sont comparés avec des méthodes de détection similaires, la stéganalyse par apprentissage supervisé est donc exclue de ce comparatif car elle repose sur une méthodologie très différente. Les quatre méthodes de stéganalyse comparées dans cette section sont : le test proposé $\hat{\delta}$, le test du RV proposé dans [4], le test du χ^2 [12] et le RS [6].

Un autre test du RVG, dont les performances ne peuvent être calculées, est présenté ; ce dernier utilise une estimation $\widehat{\Omega}_K^{\mathcal{W}}$ basée sur une décomposition en ondelettes conservant autant que possible les coefficients que pour le test proposé $\hat{\delta}$. Les résultats d'une comparaison à grande échelle sur les 1 338 images de UCID (<http://vision.cs.aston.ac.uk/datasets/UCID/>) sont présentés dans les figures 2. Dans les expérimentations, les paramètres des discontinuités ont été estimés en se basant sur la méthodologie de [1] et la variance non-stationnaire du bruit a été estimée en adaptant la méthode de [5].

La pertinence des recherches théoriques est illustrée dans la table 3b, présentant les résultats d'une simulation Monte Carlo sur l'image 3a. Les 3 canaux de l'image ont été successivement bruités (suivant une loi $\mathcal{N}(0,1)$), analysés, stéganographiés (avec un taux $R=0.1$) puis analysés une nouvelle fois. On constate que comparativement au RS, la méthode proposée est stable : les valeurs empiriques et théoriques du seuil τ_α sont en effet très proches. Par ailleurs, le modèle du contenu de l'image proposé semble pertinent puisqu'on constate qu'une approximation par shrinkage d'ondelettes offre une moindre puissance.

5 Conclusions et perspectives

Dans ce papier un premier lien a été établi entre la stéganalyse et la modélisation des images. Une description physique de l'acquisition d'une image et de la scène

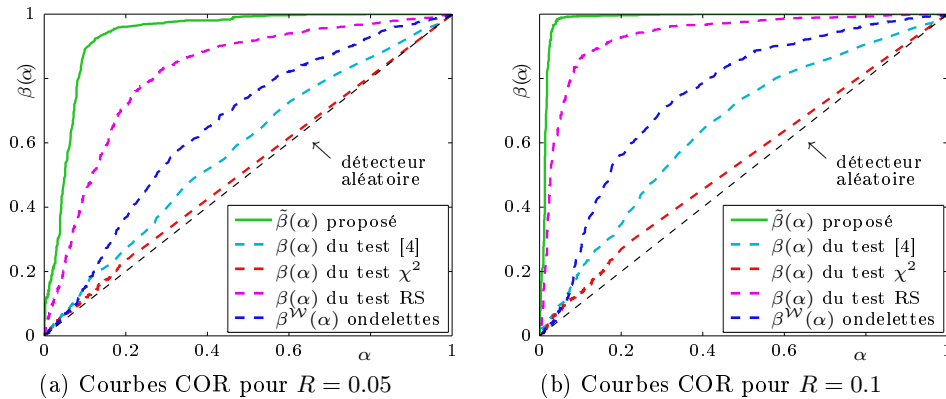
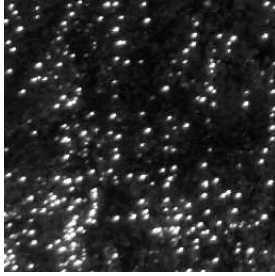


Fig. 2: Comparaison des fonctions de puissance $\beta(\alpha)$ avec la base d'images UCID.



α	Canal Rouge		Canal Vert		Canal Bleu	
	0.01	0.002	0.01	0.002	0.01	0.002
R	572	617	441	507	3346	3416
S	346	380	296	345	1952	1990
$\delta^{\mathcal{V}}$	2.872	3.573	2.857	3.480	3.153	3.842
$\hat{\delta}$	2.362	3.049	2.276	2.733	2.287	2.873
δ_{NP}^*	2.326	2.878	2.326	2.878	2.326	2.878

*valeurs de seuils théoriques calculées avec $\Phi^{-1}(1 - \alpha)$

(a) Christmas Tree

(b) Valeur du seuil empirique τ_α

Fig. 3: L'image “Christmas tree” et les seuils empiriques de détection τ_α .

a conduit à un modèle d’image ad-hoc. Une linéarisation de ce modèle permet une estimation simple et précise de l’image de couverture. En utilisant la théorie de la décision statique, le test optimal pour détecter la substitution de LSB a été explicité et permet de calculer les meilleures performances atteignable. Pour une applications pratique, un test est proposé avec des (faibles) pertes de performances quantifiées. Les résultats numériques présentés montrent que le modèle proposés est suffisamment général être appliqué à des images naturelles au contenu varié et que le détecteur proposé offre des bonnes performances.

References

1. Bruni, C. & al. : Identification of discontinuities in blurred noisy signals. Circuit and systems-I, IEEE Trans. on 44(5), 422 – 433 (May 1997)
2. Böhme, R.: Advanced Statistical Steganalysis. Springer Publishing, 1st edn. (2010)
3. Cox, I., Miller, M., Bloom, J., Fridrich, J., Kalker, T.: Digital Watermarking and Steganography. Morgan Kaufmann, 2nd edition edn. (2007)
4. Dabeer, O. & al. : Detection of hiding in the least significant bit. Signal Processing, IEEE Trans. on 52(10), 3046 – 3058 (oct 2004)
5. Foi, A. & al. : Practical poissonian-gaussian noise modeling and fitting for single-image raw-data. Image Processing, IEEE Trans. on 17(10), 1737–1754 (Oct 2008)
6. Fridrich, J., Goljan, M., Du, R.: Reliable detection of LSB steganography in color and grayscale images. IEEE Multimedia 8, 22–28 (2001)
7. Goodman, J.W.: Introduction to Fourier Optics. Roberts & Co. Publishers, 3rd edn. (2005)
8. Healey, G.E., Kondepudy, R.: Radiometric CCD camera calibration and noise estimation 16(3), 267–276 (1994)
9. Ker, A.D. & al : The square root law of steganographic capacity. In: MM&Sec ’08, 8th International Workshop. pp. 107–116 (2008)
10. Lehman, E.: Testing Statistical Hypotheses, 2nd edn.. Chapman & Hall (1986)
11. Mumford, D., Shah, J.: Optimal approximations by piecewise smooth functions and associated variational problems. Com. on Pure and Applied Math. (5), 577–685
12. Westfeld, A., Andreas, P.: Attacks on steganographic systems. In: Information Hiding, 3rd International Workshop. pp. 61–76 (1999)