# Identification de Développements d'Images par Matrices de Corrélations

Antoine MALLET[1] Rémi COGRANNE[1] Patrick BAS[2] Quentin GIBOULOT[3]

[1]Université de Technologie de Troyes, LIST3N, 12 Rue Marie Curie, CS42060, 10004 Troyes Cedex, France

[2]Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRIStAL, Bât. ESPRIT, Avenue Henri Poincaré 59655 Villeneuve d'Ascq

[3]Czech Technical University in Prague, Faculty of Electrical Engineering, Prague, Czech Republic

**Résumé –** De récents travaux d'analyse de la source d'une image naturelle, liés à des problématiques en forensique comme la stéganalyse ou la détection de *deep fake*, indiquent que la chaîne d'acquisition et de traitement impacte la nature du bruit présent dans les images résultantes. Parmi les nombreux paramètres entrant en compte dans la définition d'une source d'images, la chaîne de traitement a été identifiée comme celui ayant le plus d'impact sur la nature du bruit généré. Nous présentons dans ce travail exploratoire la preuve que l'identification du développement d'une image basée sur la corrélation de ce bruit est possible. Notre approche, non supervisée et gloutonne, présente des résultats prometteurs. Pour ce faire, nous étudions plusieurs méthodes d'estimation du bruit, ainsi que différentes mesures de similarités entre les corrélations.

**Abstract –** Recent works on the analysis of the source of a natural image, linked to forensic problems such as steganalysis or deep fake detection, show that the acquisition and processing pipeline have an impact on the nature of the noise present in the resulting images. This impact is particularly visible in the correlation of neighboring pixels. Among the numerous parameters involved in the definition of a source, the processing pipeline has been identified as the most impactful on the nature of the generated noise. We present here an exploratory study, opening a path towards an unsupervised method for image source identification, based on correlations of the heteroscedastic noise of the developped images, extracted from an noise estimation of the images. Several estimation methods are studied, aswel as several similarity measures between correlations.

## 1 Introduction

L'identification de la source d'une image consiste à reconnaître la chaîne d'acquisition et de traitement d'une image naturelle. Ce problème est devenu, en traitement d'image et notamment en forensique, un enjeu important ces dernières années. En stéganalyse par exemple, l'identification de la source d'une image permet de contrer le problème de généralisation de détecteurs à des sources inconnues à l'entraînement, baptisé *Cover-Source Mismatch*.

Pour comprendre ce phénomène, [9] définit la source d'une image comme la combinaison d'un appareil photo, de paramètres d'acquisition et de paramètres de développement. L'étude montre que les opérations de développement induisent les plus grands décalages entre les sources. Notons également [7, 11, 6], qui présentent l'impact de la source sur la corrélation des pixels voisins. Ces travaux illustrent le fait que des sources différentes peuvent engendrer des corrélations entre pixels très différentes.

Par ailleurs, comme il a été montré dans [5], l'acquisition d'une image produit de manière inhérente un bruit hétéroscédastique, i.e. le bruit associé à chaque pixel dépend de la valeur du pixel. Pour un pixel de valeur $\mu$, le bruit associé suit une loi gaussienne :

$$S \sim \mathcal{N}(0, \sqrt{a\mu + b}), \tag{1}$$

où $a$ et $b$ sont les paramètres hétéroscédastiques. Cette propriété du bruit introduit une difficulté majeure dans l'estimation de la corrélation du bruit d'une image, puisque celle-ci n'est pas identique partout dans l'image. Pour cela, [8] développe une méthode de calcul de la covariance en se basant sur l'estimation des paramètres hétéroscédastiques, et sur l'hypothèse de l'accès en mode « boîte noire » des opérations de développement. Cela permet d'estimer la matrice de covariance de chaque bloc de l'image.

Comparer la distribution du bruit de deux images par ce biais est peu aisé, car assez contraignant, notamment dans un contexte expérimental « proche » de la réalité. Dans un objectif de se rapprocher d'un contexte opérationnel, nous préférons supposer ne pas avoir accès à la chaîne de développement, mais plutôt à un ensemble d'images développées, dont nous connaissons l'étiquetage, mais pas le développement. En stéganographie toujours, [11] étudie la covariance estimée sur du bruit, s'affranchissant de la contrainte hétéroscédastique en générant un bruit gaussien centré sur une valeur de photo-site constante $\mu = 2^{12}$. L'estimation de la covariance est alors directe.

Il n'existe pas à notre connaissance de méthode de calcul en aveugle de la matrice de corrélation du bruit développé sur des images naturelles. Un tel outil serait bénéfique, car il permettrait de disposer d'une empreinte de la chaîne de développement de l'image.

Dans ce travail, principalement exploratoire, nous développons une telle méthode, présentée en section 2. De plus, comme nous le montrons dans la section 3, un algorithme d'étiquetage sans phase d'apprentissage basé sur ces empreintes offre des résultats encourageants. Nous comparons plusieurs stratégies d'extraction du bruit : notre méthode, basée sur un filtre passe-haut usuel, ainsi que *Noiseprint* basée sur des réseaux siamois [4] et *Non-Local-Mean* [2], deux algorithmes

à l'état de l'art. Nous expérimentons également plusieurs mesures de similarités entre les empreintes. Une courte synthèse et les perspectives ouvertes par ces recherches concluent notre propos en section 4.

# 2 Méthodologie

La matrice de corrélation d'une image décrit les inter-dépendances entre pixels calculées dans le domaine spatial. Calculer la matrice de corrélation d'une image de taille $\gamma \times \gamma$ devient rapidement impossible (elle est de taille $\gamma^2 \times \gamma^2$). En considérant que le canal de traitement est stationnaire, c'est à dire qu'il impacte le bruit uniformément sur toute l'image, nous pouvons toutefois calculer la matrice de corrélation. Nous définissons de plus petites portions de l'image, qui forment une série d'échantillons identiquement distribués. Ici, nous avons considéré des blocs de taille $8 \times 8$, et donc des covariances de taille $64 \times 64$.

Dans la suite, les sources sont indexées par les indices $k$ et $l$, t.q. $(k, l) \in [0, N_S]^2$, et avec $N_S$ le nombre de sources, les images par les lettres $i$ et $j$, t.q. $(i, j) \in [0, N_k]^2$, et $N_k$ le nombre d'images issues de la source $k$. Les pixels d'un bloc et les coefficients d'une matrice de corrélation sont indexés par les lettres $u$ et $v$. les blocs $B$ sont traités sous forme vectorielle, en concaténant ses lignes.

## 2.1 Principe de la méthode

Notre approche consiste à estimer la corrélation du bruit d'une image JPEG décompressée dans le domaine spatial, afin de déterminer de quelle source, en moyenne par rapport à des corrélations d'autres images, celle-ci se rapproche le plus. La Figure 1 schématise la méthode présentée. Pour calculer la corrélation du bruit, il nous faut une méthode d'estimation du bruit. En raison de son hétéroscédasticité, nous ne pouvons pas calculer une corrélation globale, calculée empiriquement sur une agrégation de zones de l'image de luminances différentes. C'est pourquoi nous proposons de d'abord filtrer l'image par un filtre passe-haut de Laplace en 4- (ou 8-) connexités, ne prenant pas (ou prenant) en compte les voisins diagonaux :

$$\mathcal{F} = I_i^{(k)} \circledast h_L, \tag{2}$$

où,

$$h_L \in \left\{ \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix}, \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix} \right\}, \tag{3}$$

ou par un filtre de Sobel du second ordre :

$$\mathcal{F} = I_i^{(k)} \circledast h_{S_v} + I_i^{(k)} \circledast h_{S_h}, \tag{4}$$

où,

$$(h_{S_v}, h_{S_h}) = \left( \begin{bmatrix} 1 & 2 & 1 \\ -2 & -4 & -2 \\ 1 & 2 & 1 \end{bmatrix}, \begin{bmatrix} 1 & -2 & 1 \\ 2 & -4 & 2 \\ 1 & -2 & 1 \end{bmatrix} \right), \tag{5}$$

et $I_i^{(k)}$ désigne la $i$-ième image de la $k$-ième source, et $\circledast$ désigne l'opérateur de convolution. Par un seuil sur la valeur de la filtrée, nous pouvons écarter les blocs contenant un contour :

$$\mathcal{B} = \{B_i^{(k)} \mid b_{i,u}^{(k)} < T, \forall\, 0 \leq u < 64, \forall\, 0 \leq i < N\} \tag{6}$$

Celui-ci permet de supprimer la majorité de la composante sémantique de l'image filtrée. En ne conservant que les blocs de l'image ne correspondant pas à un contour, nous nous assurons de prendre en compte des blocs centrés. Enfin, nous calculons la corrélation $C$ des blocs, qui normalise la covariance par la variance de chaque pixel :

$$C_i^k = (c_i^{(k)})_{u,v} = \frac{\mathrm{Cov}(b_{i,u}^{(k)}, b_{i,v}^{(k)})}{\sqrt{\mathrm{Var}(b_{i,u}^{(k)})\mathrm{Var}(b_{i,v}^{(k)})}}, \tag{7}$$

où

$$\mathrm{Cov}(b_u, b_v) = \mathbb{E}[b_u - \mathbb{E}(b_u)]\mathbb{E}[b_u - \mathbb{E}(b_v)] \tag{8}$$

La corrélation n'est donc pas exactement celle du bruit de l'image développée, mais celle de la réponse du filtre. Cette opération ne pénalise toutefois pas la détection, dans la mesure où nous ne comparerons que des corrélations de réponses de filtres identiques.

Il nous faut maintenant un outil de mesure de la similarité de deux matrices de corrélations. Deux options sont évaluées dans cet article. Tout d'abord, la distance de Frobenius, induite par la norme de Frobenius, ainsi définie :

$$\|C_i^{(k)} - C_j^{(l)}\|_F = \sqrt{\sum_{u,v} (c_{i,u,v}^{(k)} - c_{j,u,v}^{(l)})^2} \tag{9}$$

Le coefficient de corrélation de Pearson ensuite, qui pénalise des structures de corrélations différentes plutôt que des écarts d'amplitudes à l'instar de la norme de Frobenius :

$$r_{C_i^{(k)}, C_j^{(l)}} = \frac{\sum_{u,v} (c_{i,u,v}^{(k)} - \widehat{C_i^{(k)}})(c_{j,u,v}^{(l)} - \widehat{C_j^{(l)}})}{\sqrt{\sum_{u,v} (c_{i,u,v}^{(k)} - \widehat{C_i^{(k)}})^2 \sum_{u,v} (c_{j,u,v}^{(l)} - \widehat{C_j^{(l)}})^2}} \tag{10}$$

## 2.2 Détection de sources

Pour identifier la source d'une image, nous mesurons l'écart entre sa matrice de corrélation et celle de l'ensemble des images de la base de référence, qui contient les images dont on connaît les développements. Ainsi, pour une mesure de distance $d$, le label attribué à l'image test $l_t$ est :

$$l_t = \min_k \frac{1}{N_k} \sum_{i=1}^{N_k} \|C_t - C_j^{(l)}\|_F, \tag{11}$$

et dans le cas d'une mesure de corrélation pour quantifier la similarité de deux matrices, on cherche alors à maximiser (11).

# 3 Résultats expérimentaux

## 3.1 Conditions expérimentales

Nous avons considéré 9 chaînes de traitement, présentées dans la Figure 2. Nous avons fait varier l'application d'un noyau de convolution, pour rehausser les contours ou ajouter du flou à l'image, ainsi que le facteur de qualité de la compression JPEG. Ce dernier est volontairement élevé : à un facteur de qualité plus faible, le bruit tend à être écrasé, et les covariances à ne rien pouvoir décrire. Les expériences présentées sont

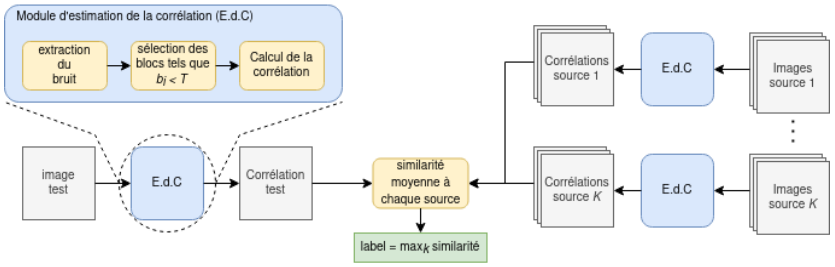


FIGURE 1 : Principe de la méthode. La similarité est mesurée soit à l'aide d'une distance, soit à l'aide d'un coefficient de corrélation.

réalisées sur des images issues de la base Alaska [3]. Afin de s'assurer que la détection de source ne soit pas faussée par le contenu des images, chaque image développée par une chaîne de traitement ne peut l'être par une autre. 500 images sont développées par chaque canal ; notre base contient donc 4500 images au total. Pour chaque expérience, 50% des données sont utilisées dans la base de référence, et 50% forment notre base de test. Chaque expérience est réalisée 5 fois, et le résultat moyen est rapporté. Nous avons fait le choix de ne considérer que le label prédit ; une autre possibilité aurait été de regarder si la source réelle de l'image figure dans le top-$x$ des sources prédites. En raison du petit nombre de sources, cette approche nous a paru peu pertinente.

Enfin, nous souhaitions comparer notre méthode avec les performances de l'algorithme état de l'art présenté dans [1] et repris par [9]. Cet algorithme utilise un ensemble de classifieurs binaires. Chaque classifieur est une méthode ensembliste, entraînée pour discriminer deux sources entre elles. A chaque paire de source correspond un classifieur ensembliste. La source prédite est alors celle recueillant le plus de vote de la part des classifieurs binaires. Ces détecteurs utilisent les caractéristiques DCTR [10], initialement imaginées pour la stéganalyse. Sur des sources proches de celles utilisées par les deux études citées, nous avons obtenu des résultats très proches de ceux présentés dans l'étude. En revanche, sur le jeux de neuf sources utilisées dans cet article, le détecteur de source offre des performances quasi parfaites, avec des taux de détection $> 99\%$. Cela s'explique vraisemblablement par la grande différence entre les sources considérées dans le présent article.

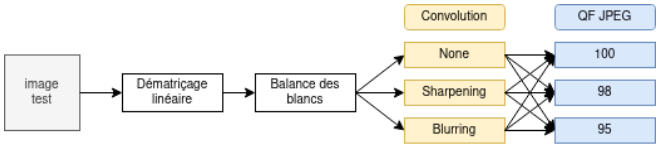


FIGURE 2 : Génération des 9 sources. 3 sources n'opèrent pas de convolution. Toutes les sources appliquent une balance des blancs.

## 3.2 Résultats

Nous commençons par une inspection visuelle, qui valide la pertinence de notre méthode. Pour une image donnée dans la Figure 3(a), nous présentons les matrices de corrélation obtenues par les 5 méthodes étudiées : Figures 3(b)-(f). Nous pouvons observer que les 3 filtres passe-haut engendrent des

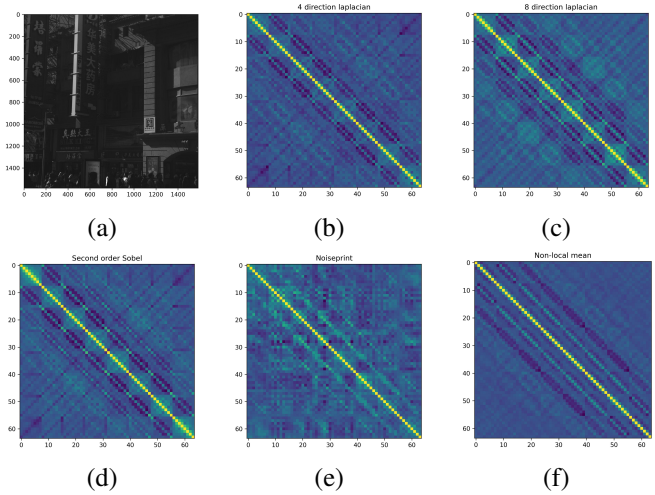


FIGURE 3 : Visualisation de la matrice de corrélations du bruit de l'image (a) après extraction par : (b) Laplace 4-connexités ; (c) Laplace 8-connexités ; (d) Sobel ; (e) Noiseprint ; (f) Non-Local Mean.

matrices relativement similaires. *Noiseprint* semble en produire des davantage bruitées, et *Non-Local Mean*, des moins bruitées. A noter que dans pour le cas des images les moins bruitées, c'est à dire celles ayant subi un flou ainsi qu'un QF $< 100$, *Non-Local Mean* est défaillant, ce qui engendre des corrélations aberrantes.

Dans le tableau 1, nous rapportons les performances de chaque extraction du bruit. Nous voyons tout de suite que la norme de Frobenius comme mesure de similarité est largement inférieure au coefficient de corrélation de Pearson en terme de performance. Cela corrobore l'idée qu'une mesure de corrélation comme mesure de similarité serait plus robuste qu'une distance, car robuste à un facteur d'échelle. On observe également la très mauvaise performance de *Non-Local Mean* en comparaison aux 4 autres méthodes. Comme indiqué précédemment, cela est dû au fait que cette méthode peine à estimer le bruit d'images fortement débruitées, par la convolution comme par le facteur de qualité JPEG.

Pour la méthode obtenant les meilleurs résultats, nous présentons dans la Figure 4 la matrice de confusion, qui illustre les sources les plus difficiles à dissocier. Nous pouvons souligner que le facteur de qualité et l'opération de convolution influent sur la proximité des sources. L'erreur a tendance à être commise vers le même facteur de qualité et vers la convolution qui induit moins de bruit.

Cette illustration nous donne une idée de l'importance de

TABLE 1 : Résultats expérimentaux. Les valeurs données sont les pourcentage de bonne classification. « lap4 » désigne le filtre de Laplace en 4-connexités, « lap8 » celui en 8-connexités, et « sobel » le filtre de Sobel, tous trois présentés en section 2.1 .

| Filtre | lap4 | lap8 | Sobel | NL-Mean | NoisePrint |
|---|---|---|---|---|---|
| Pearson | 84.1 | 82.1 | 82.1 | 31.1 | 80.6 |
| Frobenius | 36.1 | 33.5 | 28 | 14.6 | 22.3 |

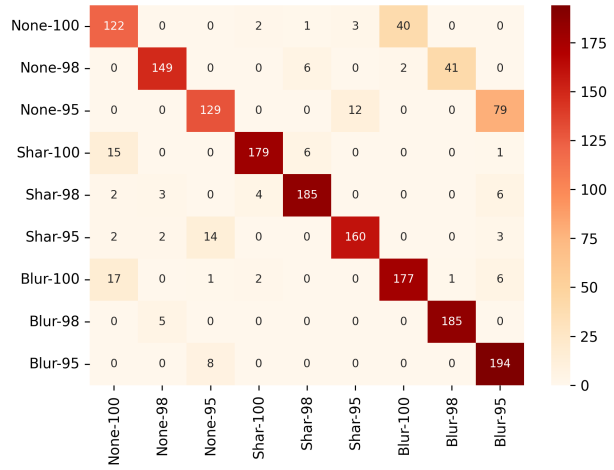


FIGURE 4 : Matrice de confusion de l'étiquetage des images en utilisant la corrélation de Pearson entre les matrices de corrélation calculées sur le filtre de Laplace en 4-connexités.

l'impact de chaque étape du canal de traitement sur la corrélation du bruit. Nous voyons que le facteur de qualité de la compression JPEG marque fortement la corrélation des pixels voisins.

# 4 Conclusion et perspectives

Dans ce travail, nous avons étudié une variété de méthodes d'estimation du bruit dans des images, afin de calculer la corrélation des pixels voisins indépendamment du contenu de l'image. Nous avons présenté une méthode d'identification de source à l'aide de différentes mesures de similarité entre les matrices de corrélations qui ne requière pas de phase d'apprentissage et qui est interprétable. Nous avons montré que notre méthode permettait d'identifier la source d'une image, avec des performances encourageantes.

Ces dernières, certes, pâlissent face à la détection quasi-parfaite de la méthode à l'état de l'art à laquelle nous nous sommes comparés. Mais nous disposons maintenant d'une empreinte interprétable de la chaîne de traitement. En mettant en lumière les difficultés parfois rencontrées par des méthodes reconnues pour estimer le bruit, nous ouvrons également une piste évidente d'amélioration. Il s'agit en particulier de mieux faire fi de la composante sémantique, par exemple en mettant en oeuvre des méthodes d'extraction du bruit plus performantes, en ajoutant des méthodes de seuillage adaptatives, ou encore en proposant des normalisations de la covariance autres que la corrélation.

# 5 Remerciements

Le travail présenté dans cet article a été rendu possible par le financement du programme de l'Union Européenne H2020, projet "UNCOVER" sous l'accord de financement No 101021687.

# 6 bibliographie

# Références

[1] D. BORGHYS, P. BAS et H. BRUYNINCKX : Facing the cover-source mismatch on jphide using training-set design. *In IHMMSec'18*. ACM, 2018.

[2] A. BUADES, B. COLL et J.-M. MOREL : A non-local algorithm for image denoising. *In 2005 IEEE CVPR'05*, 2005.

[3] R. COGRANNE, Q. GIBOULOT et P. BAS : The alaska steganalysis challenge : A first step towards steganalysis. IHMMSec'19. ACM, 2019.

[4] D. COZZOLINO et L. VERDOLIVA : Noiseprint : a cnn-based camera model fingerprint. *CoRR*, 2018.

[5] A. FOI, M. TRIMECHE, V. KATKOVNIK et K. EGIAZARIAN : Practical poissonian-gaussian noise modeling and fitting for single-image raw-data. *IEEE Trans. on Image Processing*, 2008.

[6] Q. GIBOULOT, P. BAS et R. COGRANNE : Multivariate side-informed gaussian embedding minimizing statistical detectability. *IEEE TIFS*, 2022.

[7] Q. GIBOULOT, R. COGRANNE et P. BAS : Steganalysis into the wild : How to define a source ? *In IS&T Electronic Imaging, Media Watermarking, Security, and Forensics 2018*, 2018.

[8] Q. GIBOULOT, R. COGRANNE et P. BAS : Detectability-based jpeg steganography modeling the processing pipeline : The noise-content trade-off. *IEEE TIFS*, 2021.

[9] Q. GIBOULOT, R. COGRANNE, D. BORGHYS et P. BAS : Effects and solutions of cover-source mismatch in image steganalysis. *Signal Processing : Image Communication*, 2020.

[10] V. HOLUB et J. FRIDRICH : Low-complexity features for jpeg steganalysis using undecimated dct. *IEEE TIFS*, 10:219–228, 02 2015.

[11] T. TABURET, P. BAS, W. SAWAYA et R. COGRANNE : JPEG steganography and synchronization of DCT coefficients for a given development pipeline. *CoRR*, 2020.