Stéganographie et Stéganalyse des images JPEG Couleur

Papa Ndiaye, Marc Chaumont, Mehdi Yedroudj, Ahmad Zakaria

Stéganographie et Stéganalyse des images JPEG Couleur

Papa Mamadou Ndiaye[3,4] Marc Chaumont[1,2] Mehdi Yedroudj[1] Ahmad Zakaria[1]

[1] UNIVERSITE MONTPELLIER, UMR5506-LIRMM, F-34095 Montpellier Cedex 5, France

[2] UNIVERSITE DE NIMES, F-30021 Nîmes Cedex 1, France

[3] ECOLE SUPERIEURE POLYTECHNIQUE DE DAKAR, 5005 Dakar - Fann, Sénégal

[4] CNRS, UMR5506-LIRMM, F-34392 Montpellier Cedex 5, France

{ndiaye, chaumont, yedroudj, zakaria}@lirmm.fr

Résumé

JPEG est aujourd'hui le format d'image le plus couramment utilisé pour l'échange d'images. Bien que cela en fasse un standard naturel pour la stéganographie moderne, il n'en demeure pas moins qu'il n'y a pas de contributions pour l'insertion dans des images JPEG en couleur. Les approches d'insertion considèrent en effet uniquement l'insertion dans une image JPEG en niveaux de gris, principalement parce que l'insertion dans des images en niveau de gris est déjà un problème difficile. Dans cet article, nous étudions, de manière pratique, la question de l'insertion dans une image JPEG couleur. La question principale consiste à déterminer comment doit être effectuée la répartition du message, c'est-à-dire des bits à insérer, entre les composantes de couleurs (Y, Cr, Cb) qui ont été quantifiées. Après avoir rappelé l'état de l'art, nous donnons de premiers résultats expérimentaux indiquant que l'insertion doit principalement être effectuée dans la composante de luminance.

Mots clefs

JPEG, Couleur, Stéganographie, Stéganalyse.

1 Introduction

La stéganographie est l'art de dissimuler des informations dans un support anodin et cela sans éveiller la suspicion d'une tierce personne. La stéganalyse est la discipline duale de la stéganographie et consiste à déceler une dissimulation de données dans le support. La stéganographie a été largement appliquée aux images numériques dont les images JPEG. Cependant, si d'importantes contributions ont été apportées en stéganographie et stéganalyse des images JPEG en niveaux de gris [1, 2, 3, 4, 5], rien n'existe pour autant en terme de stéganographie moderne d'image JPEG en couleur. L'insertion dans des images en couleur a été proposée récemment dans [6, 7, 8, 9]. Les auteurs proposent dans cet articles d'insérer les bits indépendamment dans chacun des canaux RGB, et ceci en coupant le message en trois parties de même taille. Ils notent que l'insertion est sous-optimale, c'est-à-dire que le résultat pourrait être plus « sûr » si l'insertion était faite en prenant en compte les trois canaux simultanément. Les auteurs remarquent également que l'insertion optimale ne donne pas les résultats escomptés en pratique.

Dans ce papier nous souhaitons également insérer dans les trois composantes (Y, Cb et Cr) d'une image JPEG mais il est évident que la proportion de bits à insérer dans chacun des canaux ne doit pas être égale. Dans ce document nous étudions donc l'impact (en termes de détectabilité par le stéganalyste) de l'insertion d'un message en fonction de la proportion insérée dans chacun des canaux. Cette étude préliminaire est basée sur une insertion via l'algorithme de l'état de l'art J-UNIWARD [1]. Cet algorithme établit une carte de coûts de détectabilité pour chacun des coefficients DCT quantifiés et utilise ensuite cette carte pour réaliser une insertion adaptative à travers l'utilisation d'un codage par STC [10]. La sécurité de nos schémas d'insertion dans des images JPEG couleur sera évaluée en extrayant des caractéristiques inspirées du Spatial-Color Rich Model [8] et en les fournissant à un classificateur d'ensemble [11]. Les caractèristiques inspirées du Spatial-Color Rich Model sont composées du SRMQ1 ainsi que de co-occurrences de résiduels obtenus à partir des composantes couleurs. Nous nous plaçons dans l'espace de couleurs YCbCr pour calculer ces caractéristiques. Nous présentons notre proposition dans la section 2 puis le protocole dans la section 3. En section 4 nous analysons et interprétons nos résultats.

2 Propositions

En confrontant plusieurs variantes du même schéma stéganographique qui différent uniquement de par la proportion du message inséré dans la luminance, nous tentons de déterminer la bonne proportion à répartir dans la luminance et les chrominances lors de l'insertion. Nous présentons dans cette section les aspects liés à l'insertion dont la stratégie de répartition du budget.

2.1 Insertion indépendante

Les approches de stéganographie couleur exploitent chaque composante de l'image en y insérant une certaine portion du message. Cette insertion peut être opérée au sein de chaque canal indépendamment des autres composantes en considérant qu'une image en couleur est une combinaison de trois images en niveaux de gris. Dans [7], Abdulrahman et al. adoptent cette stratégie de stéganographie couleur en la préférant à l'approche qui considére une unique composante (concaténation des trois canaux RGB) et en effectuant l'insertion. Cette dernière approche est en pratique plus détectable. Il pourrait également être légitime de penser qu'effectuer une synchronisation entre les canaux de couleurs permettrait d'accroître la sécurité des schémas stéganographiques. En s'inspirant des travaux dans des images en niveau de gris de Denemark et Fridrich [12], on pourrait en effet contraindre l'algorithme d'insertion à favoriser les changements de même direction sur les trois canaux YCbCr [1] . De la même façon, on pourrait s'inspirer de la synchronisation dans des images couleur RGB proposée dans l'algorithme CMD-C [13]. Malheureusement, des erreurs dans le protocole du papier CMD-C invalident les résultats obtenus ainsi que les conclusions. On peut penser que préserver la corrélation entre canaux afin d'augmenter la sécurité empirique des images couleurs pourrait être intéressant ; toutefois, la corrélation des composantes Y,Cr,Cb est très faible. La question de la synchronisation reste donc une question ouverte et elle ne sera pas traitée dans cet article.
Dans cette étude préliminaire, nous choisissons donc d'insérer indépendamment dans chacune des composantes et cela sans tenir compte des possibilités de synchronisation.

2.2 Répartition du budget

Nous souhaitons comparer la détectabilité d'une technique distribuant les bits du message sur les trois canaux couleurs Y, Cb et Cr, et celle d'une approche consistant à insérer intégralement le message dans la luminance. Nous choisissons l'algorithme J-UNIWARD 'niveau de gris' pour implémenter chacune des deux approches. On désignera par 'payload relatif' à une composante, le nombre de bits du message inséré dans une composante. Il est important de noter que les schémas stéganographiques JPEG insèrent les messages dans les coefficients DCT quantifiés. Les paramètres d'entrée de JUNIWARD sont le fichier image JPEG et un message binaire dont la taille est exprimée en bpnzac (bits par coefficient AC non nul). Pour une comparaison objective à budget constant, on n'utilisera pas l'unité bpnzac. En effet, le nombre de coefficients AC non nuls varie en fonction des images, ce qui fait que pour un même nombre de bpnzac inséré dans deux images, celles-ci contiendraient à priori des messages de tailles différentes.
Nous préférons donc travailler à budget total constant et donc utiliser l'unité bits par pixels (bpp). Ainsi, deux images stéganographiées avec le même nombre de bpp, contiendront exactement le même nombre de bits. Notons que lorsque l'insertion est habituellement effectuée en bpnzac avec des budgets entre 0.1 et 0.5 bpnzac. Pour rester dans la même gamme et donc étudier la sécurité dans une gamme de taille de message similaire, le budget doit être compris entre 0.005 bpp et 0.03 bpp. Par soucis de commodité, nous testerons 6 valeurs P de taille de message exprimés en bpp :

$$P \in \{0.005; 0.010; 0.015; 0.020; 0.025; 0.030\} \quad (1)$$

Pour chaque valeur de P, nous insérerons une proportion α dans la luminance et la proportion $\beta = \frac{1-\alpha}{2}$ dans chacune des chrominances. Nous testerons 6 valeurs différentes pour α :

$$\alpha \in \{85\%; 90\%; 93\%; 95\%97\%; 100\%\} \quad (2)$$

Pratiquement, pour un P fixé, le budget en nombre de bits est obtenu en multipliant P par le nombre de pixels de l'image. Ce budget est réparti entre les canaux de couleur en considérant que la proportion α est celle allouée à la luminance. La portion du budget restante est équitablement répartie entre les deux chrominances. On déduit ensuite le nombre de bpnzac à insérer dans chaque composante en faisant un rapport entre la portion de budget affectée à chaque composante et son nombre de coefficients AC non nuls. Cette manipulation permet d'utiliser J-UNIWARD et sa version simulée en lui passant une composante et la valeur de bpnzac.

3 Protocole Expérimental

3.1 Base d'images

Notre base d'images de couverture est construite à partir des 10000 images RAW de la BOSSBase 1.0 [6] en suivant les étapes suivantes :

— Conversion des images de la BOSSBase en 10000 images couleur 512x512 au format PPM en utilisant successivement les primitives ufraw et convert de ImageMagick. Le code source bash utilisé pour ces opérations est disponible sur `http://www.lirmm.fr/~chaumont/BOSSJPEG/macroProductPPM.sh` .

— Compression des images PPM en des images JPEG de facteur de qualité QF=75 en utilisant les primitives imwrite de Matlab.

Nous n'appliquons aucun sous-échantillonnage aux chrominances des images obtenues dans la mesure où nos recherches nous ont indiqué que le format de sous-échantillonnage le plus représentatif est le 4 :4 :4. En effet, le tableau 1 établit le format de sous échantillonnage des chrominances opéré par trois réseaux sociaux importants en mettant en relief le nombre d'images téléchargées chaque jour. Les indications sur les formats de sortie ont été obtenues en publiant une image non compressée sur chaque réseau et en la récupérant par la suite. On dénote deux fois

1. direction positive : insertion +1 ; direction négative : insertion -1

Tableau 1 – *Compression des images sur les réseaux sociaux*

Réseaux sociaux	images publiées par jour	Format
Facebook	205 millions	4 :2 :0
Twitter	342 millions	4 :2 :0
Whatsapp	700 millions	4 :4 :4

plus de trafic d'images sur Whatsapp que sur Facebook et Twitter réunis. Pour cette raison, nous choisissons de calquer notre modèle de compression JPEG sur celui de Whatsapp. Nous disposons au terme de cette phase d'une base d'images de couverture que nous nommons "Cover75444".

3.2 Elements de stéganalyse

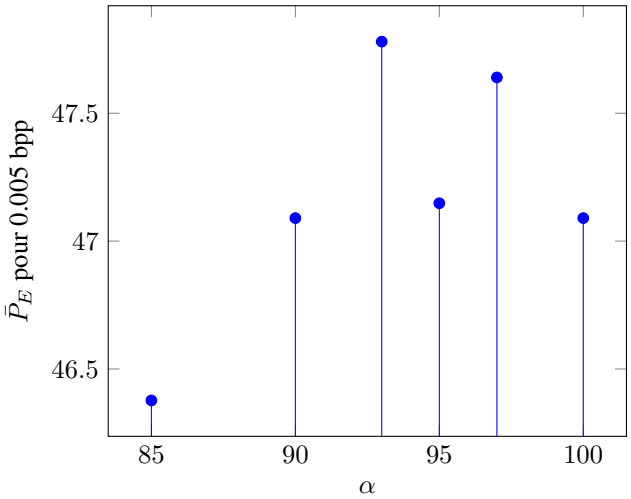
Pour évaluer les performances des schémas proposés, il est nécessaire de prendre en compte aussi bien l'aspect spatial que colorimétrique dans la phase de stéganalyse. Le Spatio-Color Rich Model (SCRMQ1) [3] a deux composantes principales qui prennent en compte cela.
La première composante est le Spatial Rich Model (SCRMQ1) [14] qui consiste en un vecteur de 12753 valeurs obtenues en calculant séparément sur chaque composante couleur des bruits résiduels via l'utilisation de nombreux filtres passe-haut, puis à quantifier les images obtenus avec un pas de quantification de 1. Ensuite les valeurs trop élevées sont tronquées (l'indice de troncature T est égal à 2). De là, des matrices de co-occurrences (horizontales et verticales) sont calculées sur chacune des images de bruits résiduels. Les trois matrices de co-occurrences sont fusionnées pour garder une même dimension que le vecteur d'origine, soit 12753 valeurs.
La seconde composante calculée dans le Color Rich Model est obtenue en nous basant sur les résiduels de chaque canal et en calculant des co-occurrences transversales. Il en résulte 5404 features (CRMQ1).
Ainsi, le Spatial Color Rich Model comprend 18157 features au total. Etant donné que la stéganographie JPEG s'opère dans un espace de couleur YCbCr, nous proposons une alternative au SCRM et que nous calculons directement à partir des composantes YCbCr. Le filtrage des résiduels ainsi que le calcul des cooccurrences s'effectuent de la même manière que pour le SCRMQ1 et les 18157 features résultantes seront appelées YCbCr-SCRM. Elles sont ensuite utilisées par le classificateur d'ensemble [11] pour la phase d'apprentissage et de test. Ce classificateur est formé par un ensemble de détecteurs binaires implémentés via le calcul des déterminants linéaires de Fisher et agissant chacun sur des portions réduites des caractéristiques de l'image. Nous utilisons le classificateur dans la version qui minimise la probabilité d'erreur de classification :

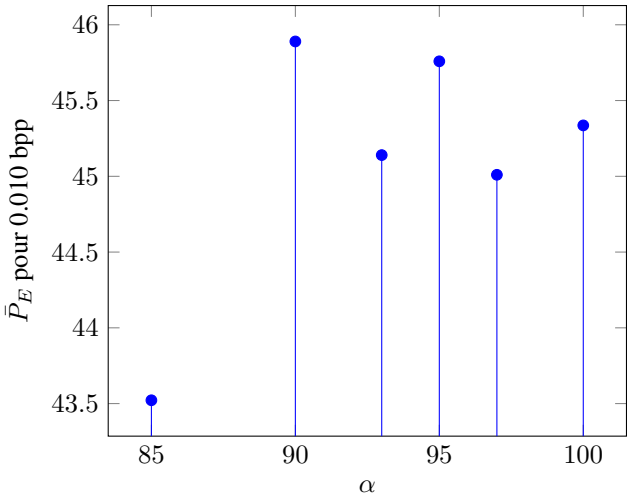
$$P_E = \min_{P_{FA}} \frac{1}{2}(P_{FA} + P_{MD})$$

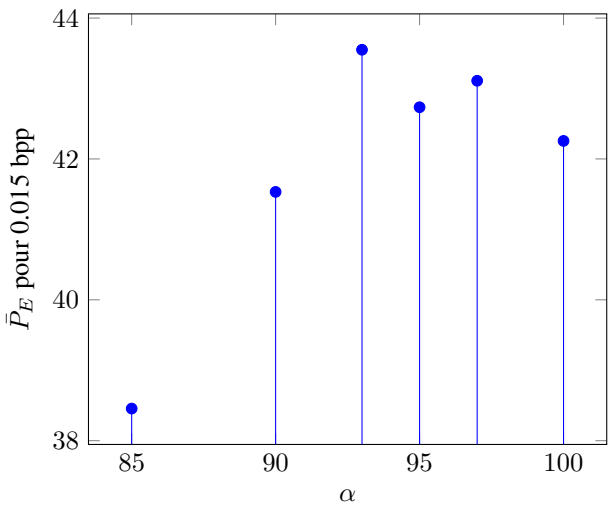
avec P_{FA} la probabilité de fausse alarme et P_{MD} la probabilité de détection ratée. On calcule la moyenne des probabilités d'erreurs P_E sur 10 différents scénarii d'apprentissage et de test au cours desquels les 10000 images de couverture et les 10000 images stéganographiées sont réparties en deux parties égales de telle sorte que chaque image stego soit associée à son image de couverture correspondante.

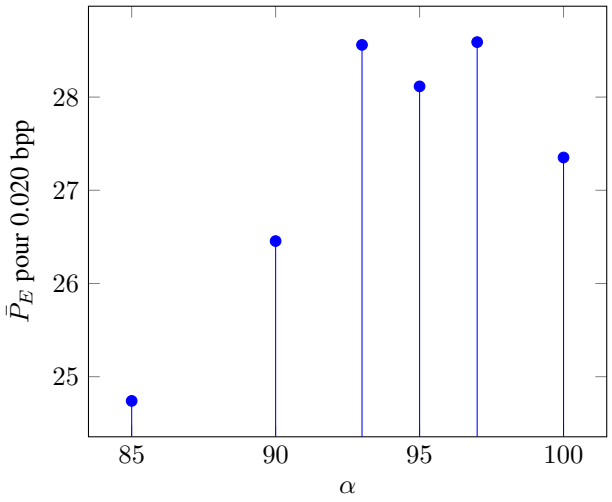
4 Résultats et discussion

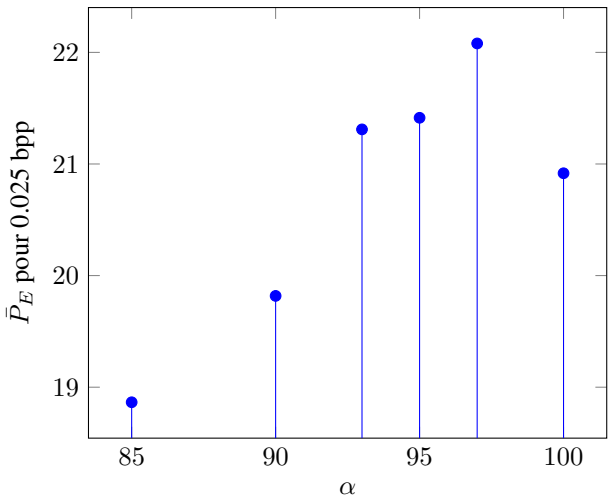
On présente dans cette partie les résultats des travaux menés en nous basant sur le protocole expérimental défini en section 2. Au total, six courbes correspondant chacune à un budget en bpp fixé (Eq. 1) et à une proportion α définie (Eq. 2). En abscisse, l'on retrouve les proportions de répartition du budget et en ordonnées la moyenne des probabilités d'erreur pour un payload particulier. Nous conviendrons de noter α* la proportion optimale de message à affecter à la luminance.











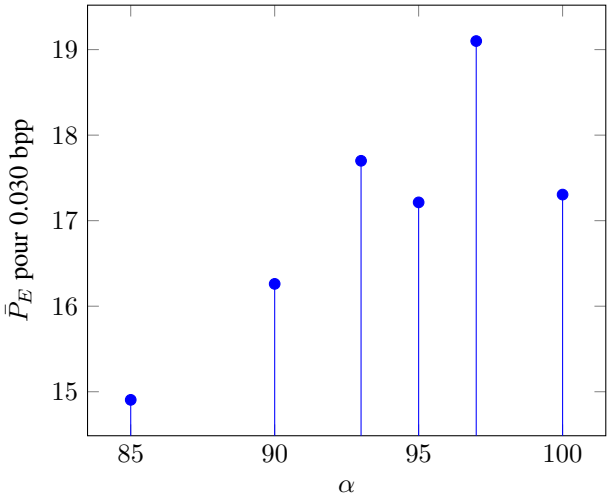


Figure 1 – *Détectabilite des variantes de JUNIWARD avec distribution du payload entre les canaux de couleur. Les courbes correspondent respectivement à une insertion de payload de 0.005 bpp, 0.010 bpp, 0.015 bpp, 0.020 bpp, 0.025 bpp et 0.030 bpp*

Nous pouvons tirer d'importantes observations des résultats présentés à la figure 1. De prime abord, les stratégies de distribution des payloads de 0.005 à 0.03 bpp montrent que l'insertion d'une portion du message dans les composantes de chrominances peut améliorer les performances du schéma stéganographique. La proportion adéquate α* à affecter à la luminance oscille entre 90% et 97%, rendant cornélien le choix de la variante de JUNIWARD à adopter. Nous observons d'autre part que les taux d'erreurs pour les payloads allant de 0.005 bpp à 0.015 bpp ne correspondent pas à des gammes d'erreur intéressantes, certaines probabilités de détection virant à l'aléatoire ($\approx 47\%$). En revanche, pour les autres payloads de 0.02 à 0.03 bpp, l'erreur de detection est maximale en 97%.
Ainsi, pour JUNIWARD à QF=75, une proportion α*=97% est un bon choix. L'insertion dans les chrominances fait donc gagner 1% en termes de sécurité. Ce gain est non négligeable (la variance étant de 10^{-6}) en stéganalyse.

5 Conclusion et perspectives

Au terme de cette étude, il apparaît qu'élaborer une stratégie de distribution du message entre les composantes couleurs dans le cadre de la stéganographie d'images JPEG couleur permet d'améliorer la sécurité du schéma d'environ 1% pour J-UNIWARD lorsque l'insertion est réalisée de maniére indépendante et à QF=75. La proportion de répartition du payload dans la luminance se situe autour de 97%. Toutefois, d'autres investigations prenant en compte d'autres facteurs de qualité, d'autres algorithmes d'insertion, d'autres bases, ainsi que d'autres approche de stéganalyse JPEG couleur nous permettront d'affiner nos conclusions.

Références

[1] Vojtech Holub et Jessica Fridrich. Digital image steganography using universal distortion. Dans *IH&MMSec 13 Proceedings of the first ACM workshop on Information hiding and multimedia security*, pages 59–68, Montpellier, France, Juin 2013.

[2] Linjie Guo, Jiangqun Ni, et Yun Qing Shi. Uniform embedding for efficient jpeg steganography. *IEEE Transactions on Information Forensics and Security*, 9(5) :814–825, Mai 2014.

[3] Xiaofeng Song, Fenlin Liu, Chunfang Yang, Xiangyang Luo, et Yi Zhang. Steganalysis of adaptive jpeg steganography using 2d gabor filters. Dans *Proceedings of the 3rd ACM Workshop on Information Hiding and Multimedia Security*, pages 15–23, Portland, Oregon, USA, 2015.

[4] Vojtech Holub et Jessica Fridrich. Low complexity features for jpeg steganalysis using undecimated dct. *Information Forensics and Security, IEEE Transactions*, 10(2) :219–228, Juin 2015.

[5] Vojtech Holub et Jessica Fridrich. Phase-aware projection model for steganalysis of jpeg images. Dans *Proceedings SPIE, Electronic Imaging, Media Watermarking, Security, and Forensics XVII*, pages 187–198, San Francisco, CA, Mars 2015.

[6] Hasan Abadulrahman, Marc Chaumont, Philippe Montesinos, et Baptiste Magnier. Color images steganalysis using rgb channel geometric transformation measures. *Wiley Journal on Security and Communication Networks (SCN) - Special Issue on Cyber Crime*, 9(15) :2945–2956, Février 2016.

[7] Hasan Abdulrahman, Marc Chaumont, Philippe Montesinos, et Baptiste Magnier. Color image steganalysis using correlations between rgb channels. Dans *Proceedings Int. Conf. Avail., Reliab., Security*, pages 448–454, Toulouse, France, Aout 2015.

[8] Miroslav Goljan, Jessica Fridrich, et Remi Cogranne. Rich model for steganalysis of color images. Dans *IEEE Int. Workshop Inf. Forensics Security*, pages 185–190, Atlanta, GA, USA, Decembre 2014.

[9] Miroslav Goljan et Jessica Fridrich. Cfa-aware features for steganalysis of color images. Dans *Proceedings SPIE 9409, Media Watermarking, Security, and Forensics*, page 94090V, San Francisco, Californie, Etats Unis, March 2015.

[10] Tomas Filler, Jan Judas, et Jessica Fridrich. Minimizing additive distortion in steganography using syndrome-trellis codes. *IEEE Transactions on Information Forensics and Security*, 6(3) :920–935, Septembre 2011.

[11] Jan Kodovsky et Jessica Fridrich. Ensemble classifiers for steganalysis of digital media. *Information Forensics and Security, IEEE Transactions*, 7(2) :432–444, Juin 2012.

[12] Tomas Denemark et Jessica Fridrich. Improving steganographic security by synchronizing the selection channel. Dans *15th ACM Workshop on Information Hiding and Multimedia Security*, pages 5–14, Portland, Oregon, USA, Juin 2015.

[13] Weixuan Tang, Bin Li, Weiqi Luo, et Jiwu Huang. Clustering steganographic modification directions for color components. *IEEE Signal Processing Letters*, 23(2) :197–201, Janvier 2016.

[14] Jessica Fridrich et Jan Kodovsky. Rich models for steganalysis of digital images. *IEEE Transactions on Information Forensics and Security*, 7(3) :868–882, Juin 2012.