



REVUE ALGERIENNE DES SCIENCES

Section -A-

- Sciences de la Nature et de la Vie

- Sciences techniques

<http://univ-eltarf.dz/fr/>

Une décomposition multi résolution pour une stéganalyse passive

Bouguerme imen^{a,*}^a Département d'informatique, Université Chadli BENDJEDID- El Tarf, Algérie.

Informations	Résumé
Mots clés : Stéganographie Stéganalyse Sélection de Variables DCT Contourlet Markov *Correspondance : b_imen88@yahoo.fr	Le but de la stéganalyse est de classer un document incriminé comme original ou comme stéganographié. Ce papier propose une méthodologie pour la stéganalyse utilisant la sélection de caractéristiques, d'ordre supérieur d'après l'image décomposé en ondelettes pour faire ressortir au mieux les propriétés et détecter la présence d'un éventuel message caché dans l'image, nous avons spécifié les images JPEG car c'est le format le plus utilisé pour la dissimulation d'information. En se basant sur ce principe, un modèle de stéganalyse universelle est proposé, utilisant les statistiques du premier ordre ainsi que l'entropie pour détecter les déviations dans l'image contenant un message dissimulé. La méthode de stéganalyse par combinaison (dct, contourlet, markov) proposée serait capable avec la présence du cover-médium de révéler si une image est stéganographiée ou non.

1. Introduction

La Stéganographie est utilisé pour cacher des messages secrets dans le milieu de couverture, et la stéganographie est une telle approche de communication secrète qu'il peut transmettre des informations sans éveiller les soupçons de l'existence de la communication secrète [1].

Le transporteur de la stéganographie peut avoir différents types de médias numériques tels que texte, image, audio, et vidéo. En raison de l'utilisation commune des images JPEG, la stéganographie a récemment attiré l'attention de plus en plus. En face de la stéganographie, Stéganalyse axée sur la découverte de la présence des données cachées, en reconnaissant ce que l'algorithme d'intégration, et d'estimer le rapport des données masquées par la suite. En général, les techniques stéganalyses peuvent être divisés en deux catégories - les approches ciblées et Stéganalyse passive.

Le premier peut également être appelé comme stéganalyse spécifique, qui est conçue pour attaquer un algorithme d'intégration spécifique connu [2].

Pour la Stéganalyse aveugle, Il a été prouvé que les images naturelles peuvent être caractérisées par des caractéristiques numériques, et les distributions des caractéristiques des images de couverture sont probablement différentes de celles de leurs images stego correspondante.

Dans cet article, nous avons étudié le concept de sténographique des données cachées dans des images numériques. Le but de cette approche est de rétablir les statistiques d'image qui sont modifiés pendant l'enrobage et sont généralement exploités par les attaques stéganographique.

Tous d'abord nous devons extraire les caractéristiques diverses d'une image JPEG, d'après le calibrage. Globalement, on observe une augmentation particulièrement forte du nombre de caractéristiques utilisées pour la stéganalyse. Dans la deuxième section, nous citons les trois domaines d'extraction des caractéristiques (caractéristique DCT, Markov, Contourlet). Dans la troisième section on termine notre article par une conclusion.

2. L'approche proposée

L'ajout d'un message à une image de couverture n'affecte pas l'aspect visuel de l'image mais peut affecter certaines statistiques. Les caractéristiques requises pour la tâche de la Stéganalyse devraient être capable d'attraper ces maladies mineures statistiques qui sont créés pendant le processus de masquage de données. Dans notre approche, nous avons d'abord vu les différentes méthodes utilisées dans l'étape d'extraction des caractéristiques [3].

2.1. Le calibrage

L'étalonnage a été introduit en 2002 comme un nouveau concept pour attaquer l'algorithme F5. Depuis lors, elle est devenue une partie essentielle d'un grand nombre de fonctionnalités basées sur la Stéganalyse passive et ciblées en format JPEG ainsi que domaine spatial. L'étalonnage a été également montré pour améliorer la précision de détection de la stéganalyse passive basée sur les caractéristiques, et consiste à estimer les caractéristiques d'image de couverture de l'image stego. Ainsi, l'effet net de l'étalonnage est de diminuer les variations d'image à image et augmenter la sensibilité des fonctionnalités à intégrer [5].

L'étalonnage est un processus utilisé pour estimer les propriétés macroscopiques de l'image de couverture et image stego. La Calibration démarre avec une image JPEG J1, L'image JPEG stego J1 est décompressé dans le domaine spatial en utilisant DCT inverse, cultivée par quatre pixels dans les deux sens, et compressé à nouveau avec la même matrice de quantification que l'image J1 stego dans les deux sens. La figure 1 montre une explication picturale de ce processus

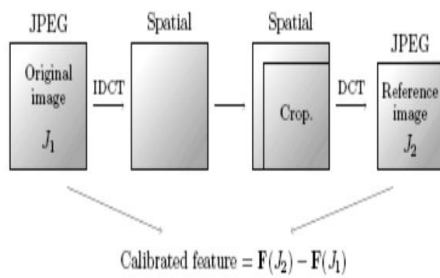


Fig. 1 : Processus de calibrage

La fonction calibrée est obtenue comme la différence entre les caractéristiques calculées pour J1 et J2 [16].

Ce travail constitue également une évaluation plus détaillée de la performance des fonctions. Ils proposent un ensemble fusionné de fonctionnalités, dont la précision de détection est remarquablement meilleure [4].

2.2. Extraction des caractéristiques DCT

Les caractéristiques d'origine DCT ont été construites par l'utilisation du 23 F fonctionnelles qui produisent un scalaire, vecteur, une matrice ou lorsqu'il est appliqué à l'image stego [14,15]. Chaque F fonctionnelle est évaluée pour l'image stego $J1$ et sa version calibrée $J2$. La fonction f calibrée est obtenue comme la différence $F(J1) - F(J2)$, si F est un scalaire, ou comme une norme $L_1 \|F(J1) - F(J2)\|_{L_1}$ si F est un vecteur ou une matrice.

Supposons que le fichier traité est une image JPEG avec une taille $M \times N$. DCT (i, j) désigne le coefficient DCT à emplacement (i, j) dans un bloc 8×8 DCT, où $1 \leq i \leq 8$ et $1 \leq j \leq 8$. Dans chaque bloc, $dct(1,1)$ est appelé le DC coefficient, qui contient une fraction importante de l'énergie sur l'image et les changements se produisent en général peu à le cours de la procédure d'intégration. Donc, il ne considère que les 63 restants coefficients AC dans chaque bloc DCT [6].

$$g_{ij}^d = \sum_{k=1}^B \delta(d, d_k(i, j)),$$

variation V :

$$V = \frac{\sum_{i,j=1}^8 \sum_{k=1}^{|I_r|-1} |d_{I_r(k)}(i, j) - d_{I_r(k+1)}(i, j)| + \sum_{i,j=1}^8 \sum_{k=1}^{|I_c|-1} |d_{I_c(k)}(i, j) - d_{I_c(k+1)}(i, j)|}{|I_r| + |I_c|}.$$

Blockiness :

$$B_\alpha = \frac{\sum_{i=1}^{\lfloor (M-1)/8 \rfloor} \sum_{j=1}^N |x_{8i,j} - x_{8i+1,j}|^\alpha + \sum_{j=1}^{\lfloor (N-1)/8 \rfloor} \sum_{i=1}^M |x_{i,8j} - x_{i,8j+1}|^\alpha}{N \lfloor (M-1)/8 \rfloor + M \lfloor (N-1)/8 \rfloor}.$$

la matrice de co-occurrence :

$$C_{st} = \frac{\sum_{k=1}^{|I_r|-1} \sum_{i,j=1}^8 \delta(s, d_{I_r(k)}(i, j)) \delta(t, d_{I_r(k+1)}(i, j)) + \sum_{k=1}^{|I_c|-1} \sum_{i,j=1}^8 \delta(s, d_{I_c(k)}(i, j)) \delta(t, d_{I_c(k+1)}(i, j))}{|I_r| + |I_c|}$$

Functional/feature name	Functional F
Global histogram	$H / \ H\ _{L_1}$
Individual histograms for 5 DCT modes	$\frac{h^{21}}{\ h^{21}\ _{L_1}}, \frac{h^{31}}{\ h^{31}\ _{L_1}}, \frac{h^{12}}{\ h^{12}\ _{L_1}}, \frac{h^{22}}{\ h^{22}\ _{L_1}}, \frac{h^{13}}{\ h^{13}\ _{L_1}}$
Dual histograms for 11 DCT values $(-5, \dots, 5)$	$\frac{g^{-5}}{\ g^{-5}\ _{L_1}}, \frac{g^{-4}}{\ g^{-4}\ _{L_1}}, \dots, \frac{g^4}{\ g^4\ _{L_1}}, \frac{g^5}{\ g^5\ _{L_1}}$
Variation	V
L_1 and L_2 blockiness	B_1, B_2
Co-occurrences	N_{00}, N_{01}, N_{11} (features, not functionals)

Tab. 1 : L'extraction des caractéristiques DCT avec 193 caractéristiques.

Afin de pallier la perte d'information due à l'utilisation de la norme L_1 et de garder la dimensionnalité des caractéristiques «raisonnable», nous avons remplacé la norme L_1 par les différences suivantes.

$$\mathbf{H}_l(J_1) - \mathbf{H}_l(J_2), l \in \{-5, \dots, 5\},$$

$$g_{ij}^d(J_1) - g_{ij}^d(J_2), (i, j) \in \{(2,1), (3,1), (4,1), (1,2), (2,2), (3,2), (1,3), (2,3), (1,4)\}.$$

$$\mathbf{C}_{st}(J_1) - \mathbf{C}_{st}(J_2), (s, t) \in [-2, +2] \times [-2, +2].$$

Après on remplace la norme L1 par les différences proposées, la dimension de l'ensemble des fonctionnalités (encore dénommée la fonction DCT jeu étendu) devient 193 [13].

2.3. Les caractéristiques de Markov

La fonction de Markov proposée en [3,7] est définie comme un modèle de différences entre les valeurs absolues des coefficients DCT voisins et un processus de Markov. Le calcul des fonctions démarre en formant la matrice $F(u, v)$ des valeurs absolues des coefficients DCT de l'image. Les coefficients DCT de $F(u, v)$ sont disposés de la même manière que pixels de l'image en remplaçant chaque bloc 8×8 pixels avec le bloc correspondant de coefficients DCT.

Ensuite, quatre tableaux de différence sont calculés selon quatre directions: horizontale, verticale, diagonale, et mineure en diagonale (en outre noté $F_h(u, v)$, $F_v(u, v)$, $F_d(u, v)$, et $F_m(u, v)$ respectivement)

A partir de ces différents tableaux, quatre matrices de probabilité de transition $\mathbf{M}_h, \mathbf{M}_v, \mathbf{M}_d, \mathbf{M}_m$ sont construites [12].

2.4. Les caractéristiques de contourlet

La contourlet transformation est une nouvelle extension à deux dimensions de la transformée en ondelettes utilisant multi directionnelle et des bancs de filtres. Pour l'extraction de caractéristiques dans le domaine discrets Contourlet Transform, ils ont avait décomposé l'image en trois niveaux pyramidale et $2n$ directions où $n = 0, 2, 4$. La figure 1 montre les niveaux et la sélection des sous-bandes pour cette décomposition. Pour la phase de décomposition pyramidale laplacien, la «Haar» filtre a été utilisé. Dans chaque échelle de grossière à fine, le nombre de directions est 1,4, et 16. Ils appliquent la décomposition pyramidale directionnelle banque de filtres et d'ignorer la sous-bande passe-bas plus beaux rapprochement, ils ont obtenu un total de 23 sous-bandes.

Et Va calculer les trois premiers moments FC normalisés pour chacune des 23

sous-bandes, ce qui donne un vecteur de caractéristiques 69-D [8,10].

2.5. Les caractéristiques fusionnées

Une combinaison directe des deux ensembles de caractéristiques produira un 517-dimension de vecteur de caractéristiques. Pour réduire la dimensionnalité de résultat, ils utilisent la moyenne $\bar{\mathbf{M}} = (\mathbf{M}_h^{(c)} + \mathbf{M}_v^{(c)} + \mathbf{M}_d^{(c)} + \mathbf{M}_m^{(c)})/4$ des quatre matrices calibrées, à la place.

Ce vecteur caractéristique a dimensionnalité 81. Ils observent que les caractéristiques de performance moyenne $\bar{\mathbf{M}}$ produit très similaire à celle de leur version

intégrale $\mathbf{M}_h^{(c)}, \mathbf{M}_v^{(c)}, \mathbf{M}_d^{(c)}, \mathbf{M}_m^{(c)}$. Après la fusion de la 193 caractéristiques DCT étendue avec les 81 en moyenne calibrée caractéristiques de Markov, la dimension de résultats caractéristiques fusionnés sera 274 caractéristiques. Le tableau 1 dresse la liste des types des caractéristiques individuelles et des symboles. Ce jeu de fonctionnalité a été choisi parce qu'il est très populaire et parce qu'elle fournit des résultats fiables stéganalyses [9,11].

3. Conclusion

Ce papier propose une méthodologie de stéganalyse par caractéristiques, pour la détection des messages cachés dans une image, en se basant sur les déviations statistiques dues à l'insertion des données dans l'image, extraites en appliquant une décomposition multi résolution sur chacune des images.

Référence :

- [1] Jessica Fridrich, " Merging Markov and DCT Features for Multi-Class JPEG Steganalysis" 2007.
- [2] J. Fridrich. " Feature-Based Stéganalyse for JPEG Images and its Implications for Future Design of Steganographic Schemes". In J. Fridrich, editor, Information Hiding, 6th International Workshop, volume 3200 of Lecture Notes in Computer Science, pages 67–81, 2005.
- [3] Y. Q. Shi, C. Chen, et W. Chen. A Markov process based approach to effective attacking JPEG steganography. In Proceedings of the 8-th Information Hiding Workshop, 2006.
- [4] J. Fridrich, T. Pevný. "Multi-class Blind Steganalysis for JPEG Images". In E. Delp and P. W. Wong, editors, Proceedings of SPIE Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents VIII, 2006.
- [5] J. Fridrich, J. Kodovský. " Calibration Revisited" . MM&Sec'09, September 7–8, 2009, Princeton,

- New Jersey, USA. Copyright 2009 ACM 978-1-60558-492-8/09/07.
- [6] J. Fridrich, M. Goljan and D. Hoge. **Steganalysis of JPEG images: "Breaking the F5 algorithm"**. In Information Hiding, 5th International Workshop, volume 2578 of Lecture Notes in Computer Science, pages 310–323, Noordwijkerhout, The Netherlands, October 7–9, 2002. Springer-Verlag, New York.
 - [7] Jon Yngve Hardeberg, Robert Jenssen. **"Image Analysis" » 16th Scandinavian Conference, SCIA 2009 Oslo, Norway, June 15-18, 2009. chapitre 33. A New Hybrid DCT and Contourlet Transform Based JPEG Image Steganalysis Technique**
 - [8] Ying Wang, **"Optimized Feature Extraction for Learning-Based Image Steganalysis"** Student Member, IEEE, and Pierre Moulin, Fellow, IEEE. 2006.
 - [9] T. Pevny and J. Fridrich. **"Towards Multi-class Blind Steganalyzer for JPEG Images"**. In Proc. IWDW, 2005, pp.39-53.
 - [10] Zhuo Li, Kuijun Lu, Xianting Zeng, Xuezheng Pan. **"A Blind Steganalytic Scheme Based on DCT and Spatial Domain for JPEG Images"**. JOURNAL OF MULTIMEDIA, VOL. 5, NO. 3, JUNE 2010
 - [11] T. Pevny et J. Fridrich. **"Determining the Stego Algorithm for JPEG Images"**. SUNY
 - [12] Binghamton, Binghamton, NY 13902_6000 April 7, 2006.
 - [13] M. Kharrazi, H. T. Sencar, and N. Memon. **"Benchmarking steganographic and steganalytic techniques"**. In E. Delp and P. W. Wong, editors, Proceedings of SPIE Electronic Imaging, Security, Steganography and Watermarking of Multimedia Contents VII, volume 5681, pages 252_263, 2005.
 - [14] Guorong Xuan¹, Yun Q. Shi², **"Steganalysis Based on Multiple Features Formed by Statistical Moments of Wavelet Characteristic Functions"**. IH 2005, LNCS 3727, pp. 262 – 277, 2005.
 - [15] Qingxiao Guan, Jing Dong, and Tieniu Tan. **"Blind Quantitative Steganalysis Based on Feature Fusion and Gradient Boosting"**, H.-J. Kim, Y. Shi, and M. Barni (Eds.): IWDW 2010, LNCS 6526, pp. 266–279, 2011.
 - [16] Pevný, T., Fridrich, J. **"Multiclass Detector of Current Steganographic Methods for JPEG Format"**. IEEE Transactions On Information Forensics And Security 3(4), 635–650 (2008)