Novel DCT Features for Detecting Spatial Embedding Algorithms

Hao Zhang^(✉), Tao Zhang, Xiaodan Hou, and Xijian Ping

Zhengzhou Information Science and Technology Institute, Zhengzhou, China
haozhang78@126.com,
{brunda, hxd2305, pingxijian}@163.com

Abstract. Traditionally, discrete cosine transform (DCT) features and Cartesian calibration are mainly utilized in joint picture expert group (JPEG) steganalysis. As well known, the steganalyzer without any modification can also work for spatial steganography. However, since JPEG and spatial embedding have different influences on DCT coefficients, this direct generalization is not quite reasonable. This essay studies the statistical moments and histogram variances of DCT coefficients under pixel domain modification. A novel steganalyzer for detecting spatial steganography is established by all alternating current (AC) coefficient features and Cartesian calibration. The experimental results state that the proposed steganalyzer outperforms the old DCT feature based method, as well as several spatial and wavelet feature based methods.

Keywords: Steganalysis · Steganography · Discrete cosine transform

1 Introduction

Image steganography is the art of secret communication in which messages are embedded in an image without drawing suspicion. There are several common choices for cover formats, mainly containing JPEG and bitmap. Secret messages are usually hidden by modifying coefficients for JPEG or by changing pixel values for bitmap.

Steganalysis, as an opposite technique, is to detect the presence of hidden data. In the earlier, Farid [1] utilized a classifier to establish a supervised learning based steganalysis framework. The study, with the help of some powerful classifier, can be converted into finding effective steganalytic features by statistical analysis. Since the embedding algorithms have different kinds of influences on cover data, a well-concerned question is how to establish a universal feature set. To make a try, Shi [2], Goljan [3], and Wang [4] extracted various features in wavelet domain, Li [5] and Pevný [6] selected several feature sets in spatial domain. A work [7] in the famous BOSS competition state that the merged DCT feature set, originally designed for JPEG steganalysis, is also useful to detect spatial steganography. Nowadays, the sets of rich model [8–10] in these three domains are brought forward to design powerful steganalyzers.

Noticing that DCT feature sets in references are designed for JPEG, we argue that they are not so appropriate for spatial algorithms. In fact, since the energy of natural image concentrates in low and middle frequency parts, the high frequency coefficient

distribute closely around zero and had better not be changed for safety. Consequently, both merged DCT features and rich model features are mostly extracted from middle and low frequency coefficients. This character makes the feature set neglect the high frequency variance in spatial embedding scheme.

In this paper, we discuss how to propose a suitable DCT feature set for detecting spatial steganography. First, the model along with the corresponding statistical analysis is given and the novel feature set is proposed in Sect. 2. Next, the experimental results are shown to verify the effectiveness of new features. Finally, the paper is concluded.

2 Model and Analysis

2.1 Model Establishment

Divide the image into blocks of size 8×8 . For any pixel block $\{c_{p,q}\}_{0 \leq p,q \leq 7}$, the stego noise $\{n_{p,q}\}_{0 \leq p,q \leq 7}$ is set to be additive and the stego block $\{s_{p,q}\}_{0 \leq p,q \leq 7}$ is obtained by the formula $s_{p,q} = c_{p,q} + n_{p,q}$. Compute the two dimensional DCT transform by

$$C_{u,v} = \frac{A_u A_v}{4} \sum_{0 \leq p,q \leq 7} c_{p,q} \cos \frac{(2p+1)u\pi}{16} \cos \frac{(2q+1)v\pi}{16}, \quad (1)$$

where $0 \leq u, v \leq 7$ and $A_u = 1/\sqrt{2}$ when $u = 0$ or otherwise $A_u = 1$. Similarly, $N_{u,v}$ and $S_{u,v}$ can be obtained and the formula $S_{u,v} = C_{u,v} + N_{u,v}$ holds.

Now we come to study the statistical difference between $C_{u,v}$ and $S_{u,v}$. Given (u, v) , let the probability density function (PDF) of $C_{u,v}$ be f_C , and that of $N_{u,v}$ be f_N . The PDF of $S_{u,v}$ can be calculated by convolution

$$f_S = f_C * f_N, \quad (2)$$

as the noise is independent from cover. Model $C_{u,v}$ by generalized Gaussian distribution. Next, to make clear the PDF of $S_{u,v}$, we need to know how $N_{u,v}$ is distributed.

Since there are many kinds of embedding algorithms, which could produce distinct influences on covers, an ideal noise model are not so easy to get. Note that the traditional work relies on Gaussian distribution, we figure that it is not quite reasonable. In much earlier research, people focus on least significant bit (LSB) replacement and LSB matching algorithms. The hypothesis that stego noises are independent from pixel to pixel was soundly suitable. Observing that $N_{u,v}$ is a weighted summation of 64 random variables, the central limit theorem guarantees the accuracy of Gaussian model. In recent years, content-based adaptive algorithms are largely used. The range or intensity of modification usually depends on some preset measurement, such as L^p norm for residues by directional filters [10], absolute values of subtractive pixels [11]. Noises among adjacent pixels could have close but complex relationship. So it should be rather difficult to select a distribution with a concrete form.

Nevertheless, we find some commonness which fits for all types of embedding schemes as far as we known. This general character can be describe by

$$f_N(x) > f_N(y), 0 < x < y \text{ or } 0 > x > y. \quad (3)$$

The random variable $N_{u,v}$ which satisfies the inequality will be single-peak distributed but not necessarily symmetrical. Compared with Gaussian distribution, the model is much more universal. In addition, we can show it really matches the generally acknowledged embedding principle.

In a spatial embedding scheme, the pixel modification is always demanded to be as smaller as possible to ensure security. As a common example, the LSB algorithms limit the alteration to the range $[-1, 1]$, and the probability of 0 is theoretically bigger than that of -1 or 1 . Even in an adaptive scheme, such as PVD [12], no matter what range the pixel difference belongs to, the alteration ranges from 0 to some fixed integer k , and the distribution also satisfies the property described by the inequality (3).

Furthermore, this property can be naturally generalized to multidimensional cases. Let $X = (x_1, x_2, \dots, x_L)$ be the vector of multiple noise variables, the density $f_N(X)$ could be larger when some x_i makes bigger in $[0, +\infty)$ or makes smaller in $(-\infty, 0]$. If we choose the 64 noises $\{n_{p,q}\}_{0 \leq p,q \leq 7}$ as the vector, then the corresponding DCT noise vector Y of $\{N_{u,v}\}_{0 \leq u,v \leq 7}$ will be obtained by a orthogonal transformation T . Note that $Y = 0$ if and only if $X = T^{-1}Y = 0$, the vector Y will be also single-peak distributed. In another word, under the transform, the property still holds for Y as well as for each component $N_{u,v}$. Therefore, this single-peak model is assumed to be appropriate for analyzing the statistical differences between f_C and f_S .

2.2 Statistical Analysis

Two classes of features are utilized here to characterize the distribution alteration, one is the histogram, the other is the PDF absolute moment. The conclusions with the associated proofs are detailed as follows.

Theorem 1. If C is generalized-Gaussian distributed and N is single-peak distributed subjected to the inequality (3), then f_C and f_S satisfy that

$$\begin{cases} f_S(0) \leq f_C(0), \\ \frac{f_S(x) + f_S(-x)}{f_S(0)} \geq \frac{f_C(x) + f_C(-x)}{f_C(0)}. \end{cases} \quad (4)$$

Proof. Consider symmetrical histogram variables

$$\begin{cases} \tilde{f}_N(x) = \frac{1}{2}(f_N(x) + f_N(-x)), \\ \tilde{f}_S(x) = \frac{1}{2}(f_S(x) + f_S(-x)). \end{cases} \quad (5)$$

Utilizing the linearity of convolution, we obtain $\tilde{f}_S = f_C * \tilde{f}_N$. Then the proof of (4) is equivalent to the proof of the equalities

$$\begin{cases} \tilde{f}_S(0) \leq \tilde{f}_C(0), \\ \frac{\tilde{f}_S(x) + \tilde{f}_S(-x)}{\tilde{f}_S(0)} \geq \frac{\tilde{f}_C(x) + \tilde{f}_C(-x)}{\tilde{f}_C(0)}. \end{cases} \quad (6)$$

These can be gotten by using the theorems in [13].

Theorem 2. Denote the mathematical expectation by $E[\cdot]$. Under the same hypothesis of theorem 1, the inequality $E[|C|^n] \leq E[|S|^n]$ always holds for each $n \geq 1$.

Proof. First, we obtain

$$E[|S|^n] = \int_{-\infty}^{+\infty} f_S(x) |x|^n dx = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_C(x-y) f_N(y) |x|^n dx dy \quad (7)$$

according to the definition of mathematical expectation. For each $n \geq 1$, the inner integral satisfies that

$$\begin{aligned} \int_{-\infty}^{+\infty} f_C(x-y) |x|^n dx &= \int_{-\infty}^{+\infty} f_C(x) |x-y|^n dx \\ &= \frac{1}{2} \int_{-\infty}^{+\infty} f_C(x) (|x-y|^n + |x+y|^n) dx \\ &\geq \frac{1}{2^n} \int_{-\infty}^{+\infty} f_C(x) (|x-y| + |x+y|)^n dx \\ &\geq \int_{-\infty}^{+\infty} f_C(x) |x|^n dx \\ &\geq E[|C|^n], \end{aligned} \quad (8)$$

where the second equality uses the symmetry of f_C and the first inequality uses the convexity of x^n ($x > 0$). Then we have

$$E[|S|^n] \geq \int_{-\infty}^{+\infty} E[|C|^n] f_N(y) dy = E[|C|^n]. \quad (9)$$

These two theorems imply that, although the embedding schemes are changeable, different stego noises invariably have the same influences on both the histogram f_C and the PDF absolute moment $E[|C|^n]$.

2.3 Proposed Feature Set

Under the guide of above results, a novel steganographic feature set can be designed. Given an image, we use non-overlapping 8×8 block DCT to get the samples of 64 coefficient random variables $\{D_{u,v}\}_{0 \leq u,v \leq 7}$. Then the integer value histograms $\{h_{u,v}(\cdot)\}_{0 \leq u,v \leq 7}$ and PDF absolute moments $\{E[|D_{u,v}|^n]\}_{0 \leq u,v \leq 7}$ can be estimated by routine calculation, including rounding and averaging.

For each coefficient $D_{u,v}$, three central values of histogram and first two PDF absolute moments constitute a useful subset:

$$\mathcal{F}_{u,v} = \left\{ h_{u,v}(0), \frac{h_{u,v}(1) + h_{u,v}(-1)}{h_{u,v}(0)}, \frac{h_{u,v}(1) + h_{u,v}(-1)}{h_{u,v}(0)}, E[|D_{u,v}|], E[|D_{u,v}|^2] \right\}. \quad (10)$$

Note that the direct current coefficient $D_{0,0}$ depends on mean value of pixel block, its distribution will be significantly influenced by cover content. So we only consider 63 AC coefficients in the basis feature set:

$$\mathcal{F} = \bigcup_{0 \leq u,v \leq 7} \mathcal{F}_{u,v} - \mathcal{F}_{0,0}. \quad (11)$$

Furthermore, the set is enhanced by so-called Cartesian calibration. The calibration technique was initially applied in JPEG steganalysis [14]. It could make the steganalyzer much more sensitive to the blockiness produced by steganography. Since the image has already been in spatial domain, we directly remove the first 4 columns and arrays of pixels. A new feature set $\tilde{\mathcal{F}}$ is then obtained by using the rest of image. Finally, the steganalyzer will be established by feature set $\tilde{\mathcal{F}} \cup \mathcal{F}$ and support vector machine (SVM) classifier with the same setup as in [6].

3 Experiment and Analysis

Ten thousand 256 gray-scale images with size 512×512 are downloaded from BOWS-2 database [15]. Then we take several algorithms containing LSBMR [16], BPCS [17], PVD, AE-LSB [11], EA [18] to produce stego images. For any algorithm, select three embedding rate (ER) between 0.1 bits per pixel (bpp) and 0.4 bpp. Exact features specified by $\mathcal{F} \cup \tilde{\mathcal{F}}$ from cover or stego image.

Each time we propose a steganalyzer, we execute a binary classification for given algorithm and ratio. That is to say, only ten thousand stego samples along with the cover samples are considered in one experiment. Half of the cover-stego sample pairs are randomly chosen for training and the left for testing.

To verify the advantage of new feature set, we would like to compare it with several other ones in references. They are two spatial feature sets named “Li110 [5]” and “Pevný686 [6]”, two wavelet feature sets named “Wang156 [4]” and “Shi78 [2]”, and two DCT feature sets named “Pevný548 [14]” and “Liu70 [19]”. They are also utilized to design six kinds of steganalyzers by SVM. The area under receiver operating characteristic curve (AUC) will be used to compare the ability of these detectors.

The results are presented in Table 1, from which we can see the proposed features always outperform the traditional merged DCT features by Pevný. This indicates that the high frequency features really contribute to detection of spatial algorithms. Note that Liu take all DCT modes as an average and the corresponding results are relatively pretty bad, we figure that this simple mixture of different frequency features degrades the classification performance. For the LSBM and BPCS algorithms, although our proposed sets are not superior to the spatial domain feature sets, they clearly outperform the wavelet domain ones.

Table 1. AUC comparison

Algorithm	ER	Li110	Pevný686	Wang156	Shi78	Pevný548	Liu72	Proposed630
LSBMR	10 %	91.41	93.49	84.62	82.48	80.50	60.85	88.05
	20 %	97.13	97.74	94.63	93.89	92.43	75.44	95.98
	30 %	98.88	98.11	97.54	97.17	96.57	82.69	97.78
BPCS	10 %	72.05	98.16	69.88	50.00	79.72	51.36	87.95
	20 %	86.38	99.67	84.96	62.98	91.58	55.81	96.38
	30 %	92.72	97.48	91.10	72.92	94.80	60.49	96.82
PVD	10 %	79.24	67.51	50.00	83.57	85.28	60.10	99.85
	20 %	92.99	78.92	65.44	90.02	92.87	70.05	99.94
	30 %	97.56	87.42	78.26	94.32	96.36	77.08	99.98
AE-LSB	10 %	62.71	67.53	50.00	68.28	77.43	56.47	95.84
	20 %	75.25	76.50	56.95	77.16	88.69	65.44	98.68
	30 %	84.10	85.78	66.14	82.90	93.80	72.97	99.51
EA	20 %	72.15	69.40	72.04	59.32	66.61	56.48	75.66
	30 %	85.74	80.58	84.49	71.97	78.87	62.03	85.23
	40 %	93.47	91.33	93.85	83.97	87.23	67.67	91.77

In addition, we are glad to show that the new feature set takes advantage in detecting edge adaptive algorithms, containing PVD, AE-LSB and EA (at 0.2bpp). Especially for PVD and AE-LSB, although the algorithms modifies less pixels than LSBMR and BPCS at the same embedding level, which reduces the effectiveness of spatial methods, they should have significantly influenced the coefficient distribution in frequency domain.

In summary, the novel feature set is well appropriate for detecting these algorithms with quite different embedding schemes. This result is consistent with the statistical analysis.

4 Conclusion

In this paper, a novel DCT feature set is proposed, and has been experimentally verified to be more effective than traditional methods, especially for detecting several edge adaptive algorithms. These results show the significance of model analysis and the effectiveness of high frequency information. However, the detection ability at low embedding rates is still not satisfying. In the future, we plan to explore better steganalyzer by using more powerful transformations.

Acknowledgments. This work was supported by the National Natural Science Foundation of China (No. 61272490). The authors would like to thank the reviewers for their insightful comments and helpful suggestions.

References

1. Farid H.: Detecting hidden messages using higher-order statistical models. In: Proceedings of the IEEE International Conference on Image Processing, pp. 905–908. IEEE Press, New York (2002)

2. Shi, Y.Q., Xuan, G., Zou, D., Gao, J.: Image steganalysis based on moments of characteristic functions using wavelet decomposition, prediction-error image, and neural network. In: *Proceedings of the IEEE International Conference on Multimedia and Expo*, pp. 269–272. IEEE Press, New York (2005)
3. Goljan, M., Fridrich, J., Holtyak, T.: New blind steganalysis and its implications. In: *Proceedings of the SPIE Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents VIII*, 607201, SPIE, Bellingham (2006)
4. Wang, Y., Moulin, P.: Optimized feature extraction for learning-based image steganalysis. *IEEE Trans. Inf. Foren. Sec.* **2**, 31–45 (2007)
5. Li, B., Huang, J.W., Shi, Y.Q.: Textural features based universal steganalysis. In: *Proceedings of the SPIE Security, Forensics, Steganography, and Watermarking of Multimedia Contents X*, 765817, SPIE, Bellingham (2008)
6. Pevný, T., Bas, P., Fridrich, J.: Steganalysis by subtractive pixel adjacency matrix. *IEEE Trans. Inf. Foren. Sec.* **5**, 215–224 (2010)
7. Fridrich, J., Kodovský, J., Holub, V., Goljan, M.: Steganalysis of content-adaptive steganography in spatial domain. In: *Filler, T., Pevný, T., Craver, S., Ker, A. (eds.) IH 2011. LNCS*, vol. 6958, pp. 102–117. Springer, Heidelberg (2011)
8. Fridrich, J., Kodovský, J.: Rich models for steganalysis of digital images. *IEEE Trans. Inf. Foren. Sec.* **7**, 868–882 (2012)
9. Kodovský, J., Fridrich, J.: Steganalysis of JPEG images using rich models. In: *Proceedings of the SPIE Media Watermarking, Security, and Forensics*, 83030A, SPIE, Bellingham (2012)
10. Holub, V., Fridrich, J.: Designing steganographic distortion using directional filters. In: *Proceedings of the IEEE International Workshop on Information Forensics and Security*, pp. 234–239. IEEE Press, New York (2012)
11. Yang, C.H., Weng, C.Y., Wang, S.J., Sun, H.M.: Adaptive data hiding in edge areas of image with spatial LSB domain systems. *IEEE Trans. Inf. Foren. Sec.* **3**, 488–497 (2008)
12. Cai, K., Li, X., Zeng, T., Yang, B., Lu, X.: Reliable histogram features for detecting LSB matching. In: *Proceedings of the IEEE Image Conference of Image Processing*, pp. 1761–1764. IEEE Press, New York (2010)
13. Kodovský, J., Pevný, T., Fridrich, J.: Modern steganalysis can detect YASS. In: *Proceedings of the SPIE Media Forensics and Security II*, 754102, SPIE, Bellingham (2010)
14. Mielikainen, J.: LSB matching revisited. *IEEE Signal Process. Lett.* **13**, 285–287 (2006)
15. Bas, P., Furon, T.: BOWS-2. <http://bows2.gipsa-lab.inpg.fr>
16. Kawaguchi, E., Eason, R.O.: Principle and applications of BPCS-steganography. In: *Proceedings of the SPIE Multimedia Systems and Applications*, 3528-464, SPIE, Bellingham (1998)
17. Wu, D.C., Tsai, W.H.: A steganographic method for images by pixel-value differencing. *Pattern Recog. Lett.* **24**, 1613–1626 (2003)
18. Luo, W.Q., Huang, F.J., Huang, J.W.: Edge adaptive image steganography based on LSB matching revisited. *IEEE Trans. Inf. Foren. Sec.* **5**, 201–214 (2010)
19. Liu, Q., Sung, A.H., Qiao, M.: Neighboring joint density-based JPEG steganalysis. *ACM TIST*. 2, Article 16 (2011)