

<http://ansinet.com/itj>

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSI*net*

Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Steganalysis of Highly Undetectable Steganography Using Convolution Filtering

¹Jiaohua Qin, ¹Xuyu Xiang, ¹Yu Deng, ²Youyun Li and ¹Lili Pan

¹College of Computer Science and Information Technology,
Central South University of Forestry and Technology, Changsha, 410004, China

²School of Traffic and Transportation,
Changsha University of Science and Technology, Changsha, 410004, China

Abstract: Highly undetectable steganography (HUGO) is one of the most advanced steganographic systems. A new methodology of steganalysis is presented against HUGO for digital images. The proposed method first obtains textural features by applying local linear transformation of convolution filtering to the image. Then, the co-occurrence matrices are constructed from horizontal and vertical direction. Finally, the ensemble classifier is used to classify. Experimental results show that the proposed steganalysis system is significantly superior to the prior arts on the detection performance and computational time.

Key words: Highly undetectable steganography, steganalysis, convolution filtering, co-occurrence matrix, ensemble classifier

INTRODUCTION

The goal of steganography is to hide the very presence of communication by embedding messages into innocuous-looking cover objects (Qin *et al.*, 2009, 2010). With the development of steganalysis, new steganography algorithms which are against steganalysis have been proposed, such as an image adaptive steganography was proposed by Hong and Chen (2012) and Sabeti *et al.* (2013). Especially some highly undetectable steganography procedures are appeared and this is a challenge to the steganalyzers. The HUGO is designed to minimize distortion to high-dimensional multivariate statistics, computed from the same pixel differences. It is currently the most secure steganography system (Pevny *et al.*, 2010a) and takes into account SPAM (Pevny *et al.*, 2010b) features which use Markov transition probabilities calculated over different images obtained by using neighboring pixels in 8 directions. First and second order transitions are taken into consideration by averaging horizontal and vertical directions yielding to one feature set and diagonal directions resulting to another one.

Fridrich *et al.* (2011a) proposed a general approach for image steganalysis which uses different domains with different modalities to combat against steganalysis aware HUGO, this method extracts 33963-dimensionality HOLMES feature, the best score they achieved on BOSSrank was 80.3%. Fridrich *et al.* (2011b) used high-dimensional image models to perform HUGO, it was tested against 1st and 2nd order SPAM, Wavelet Absolute

Moments (WAM) and Cross Domain (CDF) based steganalytic features. The feature dimensionality is 24993 and the best prediction file had a score of 80.5%. Gul and Kurugollu (2011) proposed a method to break HUGO with the detection accuracy 76.8% only.

The current trend in steganalysis is to train classifiers with increasingly more complex cover models and large data sets to obtain more accurate and robust detectors. Using high dimensional data in classification based steganalysis is problematic because of curse of dimensionality. Moreover, a practical implementation is not a trivial task as well with its high dimensionality.

This study proposes a steganalysis method against HUGO steganography. Firstly, we extracted the textural matrices by local linear transformation of convolution filtering for image. Secondly, the 22130-dimensional features are obtained by constructing the co-occurrence matrices from horizontal and vertical direction. Finally, these features are used to train ensemble classifier. The feature dimensionality in our method is reduced but the detection performance is increased.

MATERIALS AND METHODS

Local linear transform: It is generally known that information hiding leads image content imperceptible changes. The Local Linear Transform (LLT) (Unser, 1986; Xiong *et al.*, 2012) is used to extract image local neighborhood information and can capture the changes of the local stochastic textures introduced by embedding messages.

The LLT is defined as follows:

$$Y = X \odot T_M \tag{1}$$

where, X denotes an image, T_M is a nonsingular transform matrix, Y is the outputs, the “ \odot ” expresses different operation symbols. As using \odot instead of $*$, it means the image X convolves with the transform matrix T_M .

Laws local linear vectors: Since different transform matrixes T_M can extract some particular aspects of local texture property, the efficiency of the LLT analysis method depends on the choice of the transform matrix (T_M) which is equivalent to a finite impulse response filter.

Laws (1980) constructed three 1-D vectors L_3, E_3 and S_3 for image level, edge and spot texture description by convolving the vectors L_1 and L_2 (Table 1). Similarly, L_5, E_5, S_5, W_5 and R_5 are gained by convolving mutually L_3, E_3 and S_3 . The W_5 and R_5 are 1-D vectors for image wave and ripple texture. Other vectors with different sizes can be obtained analogically in this way.

Feature extraction: It is explained here how the features can be extracted to increase the detection performance of the steganalyser. Firstly, the textural residual image was obtained by local linear transformation of convolution filtering. Then 22130-dimensionality co-occurrence matrix features are constructed from horizontal and vertical direction. Feature extraction consists of following steps:

- **Step 1 (Computing textural matrices):** The textural matrices $R = r_{ij}$ are computed using convolution high-pass filters of the following form:

$$R = X * T_M \tag{2}$$

where, T_M is the laws texture template matrix from the Table 1.

- **Step 2 (Quantization):** The purpose of quantization is to make the texture feature more sensitive to embedding changes at edges and textures in the image. The quantization equation is defined as follows:

$$Q_q(r_{ij}) = \text{Floor}(r_{ij}/q) \tag{3}$$

where, q is a quantization step and floor(x) is the largest integer smaller than or equal to x.

- **Step 3 (Tuncation):** For each element $r_{ij} \in Q_q$, the tuncation processing is done to curb the textural features’ dynamic range to the (-T, T). And the truncation function is defined as follows:

$$\text{trunc}_T(r_{ij}) = \begin{cases} r_{ij} & \text{if } r_{ij} \in [-T, T] \\ T\text{sign}(r_{ij}) & \text{otherwise} \end{cases} \tag{4}$$

- **Step 4 (Co-occurrences):** The co-occurrence matrices were computed from the truncated and quantized textural matrices. The horizontal co-occurrence for $R = (r_{ij})$ is formally defined as the number of groups of three neighboring textural samples with values equal to d_1, d_2, d_3 . The operators C_v, C_h and C_m are defined analogically:

$$C_{d_1 d_2 d_3}^h(R) = \left\{ \begin{aligned} & \left[(i, j) \mid (r_{ij}, r_{i+1, j}, r_{i+2, j}) = (d_1, d_2, d_3), \right] \\ & \left[d_1, d_2, d_3 \in [-T, T] \right] \end{aligned} \right\} \tag{5}$$

- **Step 5 (Feature vectors):** The fecture vectors are co-occurrence matrices formed from neighboring textural samples and formed as follows:

$$F = (C^h(Q_q(R)), C^v(Q_q(R))) \tag{6}$$

The final features are 22130 dimensional co-occurrence features from textural matrices. The features and their parameters are shown in Table 2.

Classifier: Because of the steganalysis against HUGO using high-dimensional feature spaces and large training sets, this will greatly increase the complexity of training and classification. Ensemble classifiers offer accuracy comparable and even better to the much more complex SVMs at a fraction of the computational cost (Kodovsky *et al.*, 2012). We choose ensemble classifier in our experiments because of its efficient classification performance for large scale learning.

The ensemble classifier employs Fisher linear discriminants as base learners trained on random subspaces of the feature space. It is implemented in MATLAB as described by Kodovsky *et al.* (2012) and is available for download at <http://dde.binghamton.edu/download/ensemble/>. The optimal value of sub and L can be searched and automatic determination in ensemble classifier can be made.

Statistical analysis: The accuracy of steganalysis varies significantly across different image sources. In order to

Table 1: Laws local linear vector with different sizes

Size	Vectors		
2	$L_1 = (11)$	$L_2 = (-11)$	
3	$L_3 = L_1 * L_1 = (121)$	$E_3 = L_1 * L_2 = (-101)$	$S_3 = L_2 * L_2 = (1-21)$
5	$L_5 = L_3 * L_3 = (14641)$	$E_5 = L_3 * E_3 = (-1-2021)$	$S_5 = L_3 * S_3 = (10-201)$
	$W_5 = E_5 * S_5 = (-120-21)$	$R_5 = S_5 * S_5 = (1-46-41)$	

Table 2: Features formed by textural co-occurrence matrices and their parameters

Feature type	Parameters (s,q,m,T)	Template	Dimensionality
Fm ₃ T ₃	(3,6,3,3)	L ₃	686
		E ₃	686
		S ₃	686
Fm ₄ T ₂	(3,6,4,2)	L ₃	1250
		E ₃	1250
		S ₃	1250
Fm ₅ T ₁	(3,6,5,1)	L ₃	486
		E ₃	486
		S ₃	486
Fm ₆ T ₁	(3,6,6,1)	L ₃	1458
		E ₃	1458
		S ₃	1458
Fm ₆ T ₃	(5,6,3,3)	L ₅	686
		E ₅	686
		S ₅	686
		W ₅	686
		R ₅	686
Fm ₄ T ₁	(5,6,4,1)	L ₅	162
		E ₅	162
		S ₅	162
		W ₅	162
		R ₅	162
Fm ₄ T ₂	(5,6,4,2)	L ₅	1250
		E ₅	1250
		S ₅	1250
		W ₅	1250
		R ₅	1250

q>0 is a quantization step, m: Order of the co-occurrence, T: Threshold of truncation function. s: Span of the difference used to compute the laws texture template matrix. The equation of dimensionality is $2(2T+1)^m$

assess the proposed method and compare it to prior art under different conditions, we measured their accuracy on the following four databases:

- BOSSrank provides 1000 images (518 cover and 482 stego) for the testing
- BOSSBase ver. 0.92 contains 9074 grayscale cover as well as 0.4 rate embedded stego images with size 512×512 and downloaded from <http://boss.gipsa-lab.grenoble-inp.fr/>
- BOWS2 contains approximately 10700 grayscale images with fixed size 512×512 come from rescaled and cropped natural images of various sizes. This database was used during the BOWS2 contest and download from <http://exile.felk.cvut.cz/boss/BOSSFfinal/>
- NRCS consists of 3185 raw scans of TIFF files with size 2100×1500 or 1500×2100 and download from NRCS photo gallery at <http://photogallery.nrcs.usda.gov>. For testing, these images were resampled to 640×418 and converted to grayscale
- UCID contains 1338 raw scans of TIFF images with size 512×384 or 384×512 and download from <http://vision.cs.aston.ac.uk/datasets/UCID/ucid.html>. To preserve the original statistical structure, we use 3 color components and their average as 4 different grayscale images directly. In this way, we can obtain 5352 images. These images were converted to grayscale
- JOINT contains 7500 images. From four databases above BOSSBase Ver.0.92, BOWS2, NRCS and UCID

randomly chosen 3000, 3000, 500 and 1000 images, respectively

All of above images in databases were utilized as covers to generate stego images with HUGO. Among the stego images, images of four embedding rates 0.1, 0.2, 0.3, 0.4 bpp, are included, respectively.

RESULTS AND DISCUSSION

In the first experiment, we randomly select 50% original images and their corresponding stego-images from BOSSBase ver. 0.92 for training. The BOSSrank database is used for testing. We first compare the detection performance of our method with the prior art, Breaking HUGO. The comparison of the detection accuracy is shown in Fig. 1. In the present experiments, ensemble classifier was used with the selected parameters and with the extracted features for 10 random training and testing. The average detection performance that we obtained is as the final result of the experiments. From Fig. 1, it can be see in that the detection accuracy of our method is better than the Breaking HUGO method at the different embedding rates. And it also keeps the higher detection accuracy when the embedding rate is 0.2 bpp.

In the second experiment, we randomly select 50% original images and their corresponding stego-images from JOINT database for training. We randomly select 5000 original images and their corresponding stego-images from four databases (BOSSbase, BOWS2, NRCS and UCID) for testing. Experiment 2 was designed to

evaluate the efficiency of different image databases. Figure 2 shows the detection error P_E for four image databases. From Fig. 2, it can be seen that the BOSSbase keeps the lower detection error than the other databases.

The Receiver Operation Characteristic (ROC) curve was chosen to show the detection probability in terms of

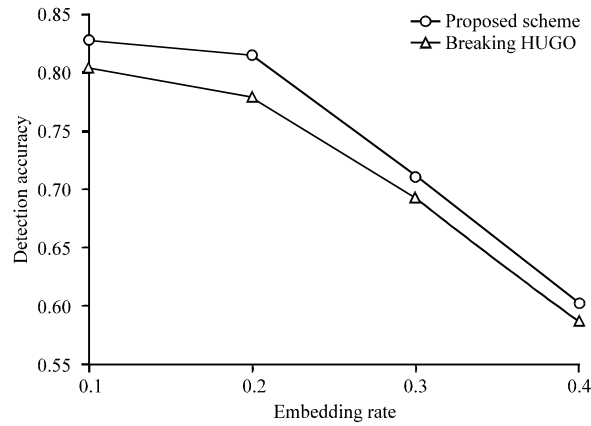


Fig. 1: Comparison of the detection accuracy with prior art: Breaking HUGO

the false positive probability. In order to better evaluate the detection performance, the ROC curves for the different image databases with different embedding rates are shown in Fig. 3a-d, where the four different curves from the left to right and from the top to the bottom stand for the image databases BOSSbase ver.0.92, BOWS2, NRCS and UCID, respectively.

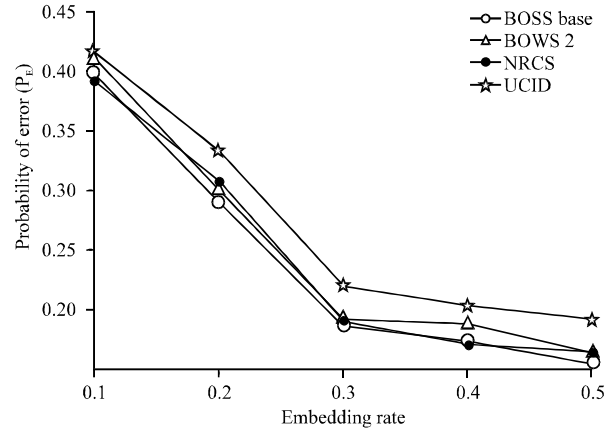


Fig. 2: Detection error P_E for four different image databases

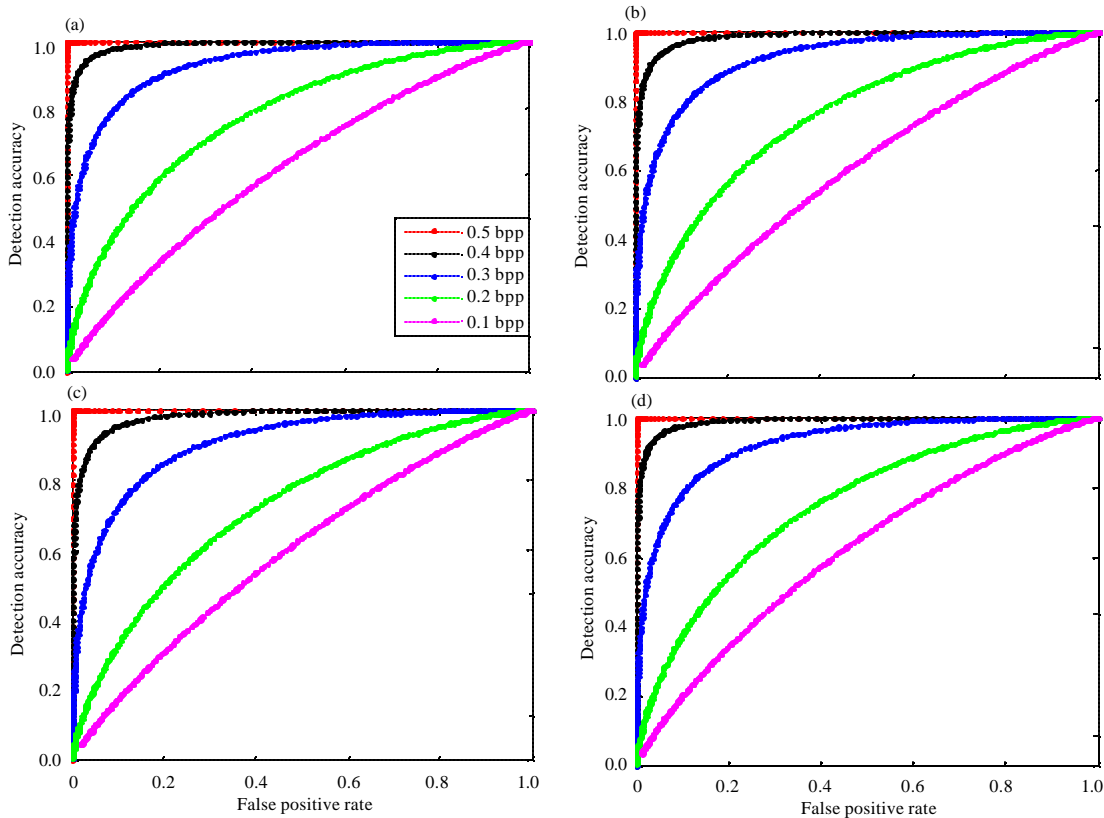


Fig. 3(a-d): ROC curves of the proposed method for four databases (a) BOSSbase ver.0.92, (b) BOSS2, (c) NRCS and (d) UCID with different embedding rates

From the Fig. 3, it can be seen that the detection accuracy of BOSSbase ver.0.92 is the highest, followed by UCID, BOWS2 and NRCS. It also shows that the BOSSBase database is better for steganography and steganalysis experimental research.

The present study is obtained textural residual image by local linear transformation, to get 22130 dimensionality co-occurrence matrix features. However, Fridrich *et al.* (2011b) extracted 24993 dimensionality features. When the embedding rate is 0.4 bit per pixel, a detection rate of 82.71% was obtained on average using BOSSRank Image Sets which is better than 80.3% by Fridrich *et al.* (2011a), 76.8% by Gul and Kurugollu (2011), also in less than 0.4 bpp, the detection effect also improved. Experimental results show that the proposed steganalysis system is significantly superior to the prior arts such as Fridrich *et al.* (2011b) on the detection performance and reduces the time complexity.

CONCLUSION

In this study, a new steganalysis method against HUGO steganography is proposed. Firstly, by local linear transformation of convolution filtering for image, the textural matrices are obtained. Secondly, the co-occurrence matrices are constructed from horizontal and vertical direction to get the 22130-dimensional features. Finally features are used to train ensemble classifier. Experimental results demonstrate that the proposed method is efficient to detect the HUGO steganography. The proposed method directly uses the convolution operation, without the residual function and greatly reduces the time complexity. The feature dimensionality is reduced but the detection performance is increased.

ACKNOWLEDGMENT

This project is supported by the National Natural Science Foundation of China (No. 61202496, 11072041, 61304208), Hunan Provincial Natural Science Foundation of China (No. 13JJ2031, 13JJ4087), Science and Technology Program of Hunan Province (No. 2014SK2025).

REFERENCES

Fridrich, J., J. Kodovsky, V. Holub and M. Goljan, 2011a. Steganalysis of content-adaptive steganography in spatial domain. Proceedings of the 13th International Conference on Information Hiding, May 18-20, 2011, Prague, Czech Republic, pp: 102-117.

Fridrich, J., J. Kodovsky, V. Holub and M. Goljan, 2011b. Breaking HUGO-the process discovery. Proceedings of the 13th International Conference on Information Hiding, May 18-20, 2011, Prague, Czech Republic, pp: 85-101.

Gul, G. and K. Kurugollu, 2011. A new methodology in steganalysis: Breaking Highly Undetectable Steganography (HUGO). Proceedings of the 13th International Conference on Information Hiding, May 18-20, 2011, Prague, Czech Republic, pp: 71-84.

Hong, W. and T.S. Chen, 2012. A novel data embedding method using adaptive pixel pair matching. Trans. Inform. Forensics Secur., 7: 176-184.

Kodovsky, J., J. Fridrich and V. Holub, 2012. Ensemble classifiers for steganalysis of digital media. Trans. Inform. Forensics Secur., 7: 432-444.

Laws, K.I., 1980. Textured image segmentation. Ph.D. Thesis, Rept. 940, Image Processing Institute, University of Southern California, Jan 1980.

Pevny, T., P. Bas and J. Fridrich, 2010a. Steganalysis by subtractive pixel adjacency matrix. IEEE Trans. Inform. Forensics Secur., 5: 215-224.

Pevny, T., T. Filler and P. Bas, 2010b. Using high-dimensional image models to perform highly undetectable steganography. Proceedings of the 12th International Conference on Information Hiding, June 28-30, 2010, Calgary, AB, Canada, pp: 161-177.

Qin, J., X. Sun, X. Xiang and Z. Xia, 2009. Steganalysis based on difference statistics for LSB matching steganography. Inform. Technol. J., 8: 1281-1286.

Qin, J., X. Xiang and M.X. Wang, 2010. A review on detection of LSB matching steganography. Inform. Technol. J., 9: 1725-1738.

Sabeti, V., S. Samavi and S. Shirani, 2013. An adaptive LSB matching steganography based on octonary complexity measure. Multimedia Tools Applic., 64: 777-793.

Unser, M., 1986. Local linear transforms for texture measurements. Signal Process, 11: 61-79.

Xiong, G., X. Ping, T. Zhang and X. Hou, 2012. Image textural features for steganalysis of spatial domain steganography. J. Electron. Imaging, Vol. 21. 10.1117/1.JEI.21.3.033015.