AN EFFECTIVE STEGANALYSIS FOR ROBUST STEGANOGRAPHY WITH REPETITIVE JPEG COMPRESSION

Jinliu Feng, Yaofei Wang, Kejiang Chen, Weiming Zhang, Nenghai Yu*

School of Cyber Science and Technology, University of Science and Technology of China
Anhui Province Key Laboratory of Cyberspace Security Situation Awareness and Evaluation

ABSTRACT

With the development of social networks, traditional covert communication requires more consideration of lossy processes of Social Network Platforms (SNPs), which is called robust steganography. Since JPEG compression is a universal processing of SNPs, a method using repeated JPEG compression to fit transport channel matching is recently proposed and shows strong compression-resist performance. However, the repeated JPEG compression will inevitably introduce other artifacts into the stego image. Using only traditional steganalysis methods does not work well towards such robust steganography under low payload. In this paper, we propose a simple and effective method to detect the mentioned steganography by chasing both steganographic perturbations as well as continuous compression artifacts. We introduce compression-forensic features as a complement to steganalysis features, and then use the ensemble classifier for detection. Experiments demonstrate that this method owns a similar and better performance with respect to both traditional and neural-network-based steganalysis.

Index Terms— Steganalysis, Robust Steganography, Repetitive Compression, Feature Combination

1. INTRODUCTION

Steganography is the art of covert communication, which hides secret message into innocent-look objects, such as texts, images, videos. Minimal distortion steganography cooperating with Syndrome Trellis Codes (STC) [1] is the mainstream steganography. However, the lossy operation of Social Network Platforms (SNPs) will invalidate the message extraction process [2], due to the strict constraint of syndrome function in minimal distortion steganography. JPEG compression is a universal operation of SNPs, so there are many robust steganography methods [2] with respect to JPEG compression being proposed. Zhang *et al.* proposed a JPEG compression resistant adaptive steganography combined with robust watermarking algorithm based on a framework of "Compression-resistant Domain Constructing + RS + STCs Codes" [3]. And their other work DMAS [4] further improved the resistance to JPEG compression. Yu *et al.* proposed GMAS [5] based on DMAS by replacing symmetric distortion with asymmetric distortion, combining with ternary STCs [1] and expanding the embedding domain to achieve strong robustness. Besides, Zhao *et al.* proposed an image preprocessing method [6] named Transport Channel Match (TCM) to resist JPEG compression. Although it is time-consuming to adjust cover images to meet the requirement of SNPs before embedding, this method can achieve considerable robustness and undetectability.

Image steganalysis focuses on whether secret message exists in digit images. As for traditional steganalysis built on manual features, which are constructed by assembling a rich model of many diverse submodels formed by image noise residual. For example, DCTR [7], GFR [8] and PHARM [9] exhibit better performances at the cost of higher dimensionality. As for deep steganalysis, such as Yedroudj-Net[10], SRNet [11] etc, show superiority to traditional manual feature sets. Although these steganalysis networks have a good detection effect in the face of general steganography, they are easily being attacked by adversarial steganography [12]. Besides, these steganalysis methods do not take into account the property of robust steganography, and the detection performance is limited, especially under low payload.

Robust steganography is very meaningful to bring theoretical experiments to the real world, so a steganalysis for robust steganography is necessary. Due TCM [6] in robust steganography is a typical class of methods using repetitive compression that can be used in a large number of ways like [13] and [14], it makes sense to do steganalysis for this case. In this paper, we propose a novel steganalysis method towards robust steganography that based Transport Channel Match by chasing both steganographic perturbations as well as continuous compression artifacts. Actually, the continuous compression artifacts have been well explored in image forensics. Inspired by the work on detecting single or double compression, we introduce error based statistical features

This work was supported in part by the Natural Science Foundation of China under Grant 62102386, 62002334, 62072421, and 62121002, and by China Postdoctoral Science Foundation under Grant 2021M693091, and by Open Fund of Anhui Province Key Laboratory of Cyberspace Security Situation Awareness and Evaluation.

E-mail: chenkj@ustc.edu.cn.

* Corresponding author.

(EBSF) [15], which are simple and effective, extracted from rounding and truncation error blocks to help improve the detection of steganalysis. We combined the compression feature and steganalysis feature with a scale. The experimental results show the combined feature performs better in detecting TCM with respect to steganalysis feature alone as well as neural-network-based steganalysis.

2. PRIOR WORK

2.1. Robust Steganography With Transport Channel Matching

In [6], Zhao *et al.* proposed an effective robust adaptive steganogaphic algorithm based on the Transport Channel Matching (TCM), in which the image is re-compressed several times with the same quantization table to match the transport channel. The diagram of TCM is shown in Figure 1, where the input are image x and parameters of transport channel, like image threshold size and quality factor of transport channel. The output is the image that matched the channel. The algorithm will first match the size of image to the threshold size of channel. Then re-compress the image to target quality factor as channel. D_1 and D_2 represent the numbers of quantization errors of two consecutive re-compressions. If D_2 is 0, or the coefficient change caused by two consecutive compressions is not much different, that is, $D_1/D_2 > 0.98$, this iteration will be returned. Otherwise, the image will be compressed again in the next iteration.

Based on the TCM algorithm, Zhao *et al.* proposed two effective methods called JCRIS and JCRISBE. As for JCRIS, secret messages are embedded to cover after TCM operation. And then images are compressed by JPEG compression. If message can not be extracted correctly, the stego images will be used to execute TCM and the new iteration restarts. Finally, most stego images are resistant to the JPEG compression from the transport channel and the message can be extracted. However, JPEG compression for some images will bring in a small number of persistent noises that cannot be effectively eliminated. Accordingly, they propose JCRISBE to deal with this problem. The secret message is encoded by BCH code and embedded into cover after TCM. If messages can not be extracted correctly after JPEG compression, they enlarge the error correction capability of BCH code and try again until the secret message can be completely extracted.

Although the robust steganography based on TCM has a better robustness and security performance compared to other robust steganography methods, it inevitably requires repeated compression of the image in the preprocessing. This will introduce other artifacts in the stego image that remain partially present after repeated compression.

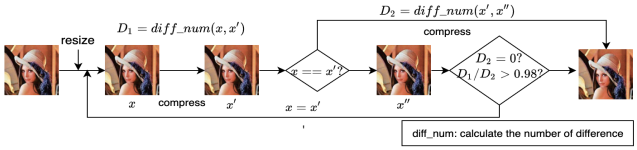

Fig. 1. The overview of TCM.

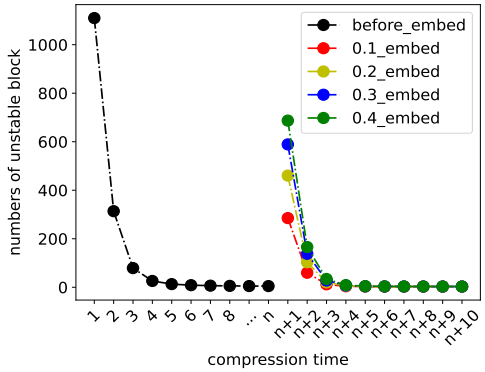

Fig. 2. The number of unstable block under different times of JPEG compression.

2.2. Forensic Method For Detecting Double JPEG Compression With the Same Quantization Matrix

In this paper, we think the detection of double JPEG compression with the same quantization matrix can help to improve the steganalysis of the robust steganography. There are some successful approaches have been presented [16, 15]. Huang's method detects single or double compression with the same quality factor according to the decreasing trend of the number of different JPEG coefficients in [16]. Yang *et al.* found truncation and rounding error block show statistical discrepancy between single and double compression in [15]. They proposed the error based statistical features (EBSF) extracted from rounding and truncation error blocks respectively. These methods distinguish different artifacts introduced during different times in JPEG compression.

In detecting single and double compression work, distinguishing single compression from multiple times compression can be easier compared to double compression. As the compression time increase, detecting accuracy increases. Therefore, the method based on compression feature can be applied to detect the stego generated by TCM.

3. PROPOSED METHOD

When the repeatedly compressed picture continues to be compressed, the compression feature changes little, and this can be well detected by forensic methods, such as EBSF [15]. But if the secret message is embedded, the compression feature will be destroyed abruptly and hard to be used by the forensic methods. Figure 2 shows the change numbers of unstable blocks under different numbers of compression times before

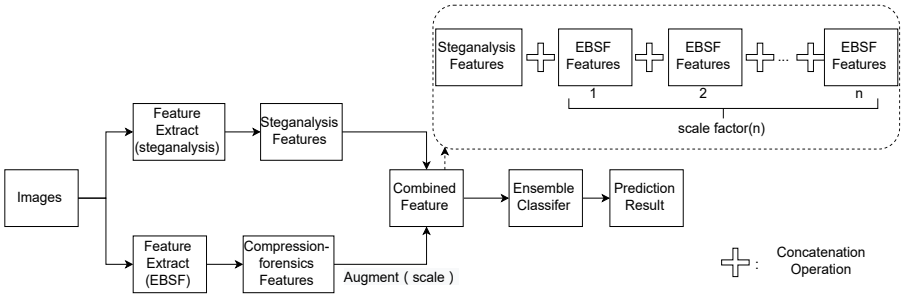

Fig. 3. The overview of the proposed robust steganalysis framework.

and after embedding with J-UNIWARD [17] under different payloads. When the payload increases, more stable blocks are destroyed and become unstable. Distinguishing the original unstable block from the embedded unstable block under high payload, EBSF features are less effective. However, it is known that steganalysis features become better when the payload is higher. On the contrary, when the payload is low, stego images are increasingly closer to cover images, and the extraction effect of steganalysis features becomes worse, but EBSF features will play a better role.

Based on the above analysis, we propose a general feature fusion framework, which combines steganalysis feature and compression feature. Figure 3 shows the framework of our approach. First, we use the traditional steganalysis feature extractor and the compression feature extractor to extract features from the cover and the stego respectively. Then, we concatenate the two different features together with a scale. After extraction and concatenation, the combination features are put into the ensemble classifier [18] for training and testing. Ensemble classifier extracts subspace from features for classification through multiple iterations. And the final classification results are output by statistical classification voting.

In this paper, we choose DCTR [7]and GFR [8] to extract steganalysis feature, and error based statistical features (EBSF) [15] as compression feature. DCTR and GFR mainly take two steps to extract handcrafted features: calculating residual maps and extracting statistical features. To highlight subtle steganographic signals, it utilizes different filter banks to calculate residuals. The EBSF features consist of three subsets [15]. The first subset is extracted directly from the pixels with four features, which contains the means and variances of absolute error values over the rounding and truncation error blocks. The second subset is calculated on DC coefficients and AC coefficients in a similar way as the first subset and with eight features. Besides, the ratio of unstable rounding error blocks constitutes the third subset and form 13-D features with the above two subsets. However, DCTR is 8000 dimensions and GFR is 17000 dimensions, and EBSF has only 13 dimensions by comparison. Obviously, concatenating features directly will make EBSF features submerged by steganalysis feature and worth nothing. In order to increase the weight of EBSF, we must expand the dimensions of EBSF. As shown in Figure 3, we replicate the EBSF features for n times and concatenate together with steganalysis feature. The scale factor is determined by experiments.

4. EXPERIMENTS

In our experiments, we use two datasets BOSSbase 1.01 [19] and BOWS2 [20]. Cover JPEG images are obtained in Matlab using the command $imwrite$. Stego is generated by Zhao's method [6]. Different steganalysis feature sets like DCTR [7], GFR [8] is chosen to extract steganalysis feature and EBSF [15] is chosen to extract compression feature. Ensemble classifier [18] is used to do training and testing. Different classifiers like Low-complexity Linear Classifier(LCLC) [21] can achieve similar results. The detection performance is evaluated by error rate P_{E} = (P_{FA} + P_{MD})/2, which means average of the probabilities of the false alarm and the missed detection over ten times.

4.1. Scale Selection

The scale factor n is a multiple of the EBSF features to be extended as shown in Figure 3. In the factor selection process, BOWS2 [20] was used as the dataset. We set the payload 0.4 bpnzac (bits per non-zero AC coefficient) at QF=75, select J-UNIWARD [17] as the distortion function to calculate costs, STC [1] as encode method and DCTR as the steganalysis feature. We set $n = 1, 10, 100$, and 1000 according to a certain scale. As shown in Figure 4, the different color lines represent P_{E} with different scaling factors, note that the black line represents the detection error rate without the EBSF combination. The experimental results show that $n = 100$ performs best when combined with EBSF to avoid being overwhelmed by the large dimensional steganalysis feature. And we will select it in the following experiments.

4.2. Comparison With Other Methods

We compare the proposed method with other methods on combination of BOSSbase 1.01 [19] and Bows2 [20]. For the limitation of computing resource, the images are resized to 256×256 with the matlab function $imresize$.

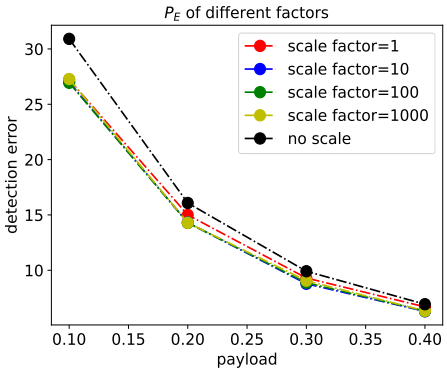

Fig. 4. Expansion factor selection.

Table 1. The comparison of detection error P_E for JCRIS based on J-UNIWARD under QF 75.

Steganalysis	0.1	0.4
DCTR	42.62%	14.72%
DCTR+EBSF	29.33%	12.09%
GFR	38.24%	12.27%
GFR+EBSF	27.81%	9.61%
EBSF	41.34%	44.35%

Table 2. The comparison of detection error P_E for JCRISBE based on J-UNIWARD under QF 75 and 85.

	75		85	
	0.1	0.4	0.1	0.4
DCTR	42.78%	17.55%	37.32%	19.21%
DCTR+EBSF	32.91%	15.47%	28.51%	17.11%
GFR	38.31%	15.75%	34.36%	13.52%
GFR+EBSF	31.41%	11.18%	27.11%	12.67%
EBSF	44.50%	46.03%	24.35%	22.99%

Table 3. The comparison of detection error P_E for SRNet and proposed method using J-UNIWARD.

Method	QF	payload	GFR+EBSF	SRNet
JCRIS	75	0.2	20.05%	19.96%
		0.4	10.28%	10.08%
	85	0.2	16.62%	18.23%
		0.4	8.9%	9.59%
JCRISBE	75	0.2	25.63%	28.78%
		0.4	11.18%	12.10%
	85	0.2	23.51%	28.78%
		0.4	12.67%	13.12%

J-UNIWARD [17] is used to calculate costs and STC [1] encode are performed in these images with different payloads from 0.1 to 0.4 bpnzac. Quality factors 75 and 85 are utilized. Table 1 and Table 2 show the improvements of our method for JCRIS and JCRISBE. As we can see JCRISBE shows superiority to JCRIS in terms of security and robustness in [6]. So we mainly show the detection results for JCRISBE. We use +EBSF as a suffix to represent features combined with EBSF.

The experiments show that the results of the combination feature have the same trend as the steganalysis feature. The detection error rate decreases as the payload increases. When combined with EBSF, there is a significant decrease in the detection error rate regardless of the steganalysis feature. Especially under low payload, the effect of combination feature improves by between 7-13%, which is a substantial improvement compared to the results of using steganalysis feature alone. The higher the error detection rate using steganalysis features alone, the greater the degree of boosting into combination features. However, the extent of the boost tends to decrease as the payload increases. This is because when the payload is high, the modification caused by embedding affects the unstable blocks. But there is still a boost of more than 1% under high payload.

In addition, separate EBSF features were performed as ablation studies. The results using EBSF alone under high quality factors were lower than those under low quality factors. This is because the artifacts generated by compression are easier to detect under high quality factors. In most cases, the detection error rate of EBSF features was higher than the others. The results prove that the facilitation effect is not only brought by EBSF features. We can conclude that the combined features can take advantage of the steganalysis feature and compression feature. We also compared our results with SRNet [11]. The experiments for SRNet were performed on PyTorch. Dataset is split into 14000 for training, 1000 for validation, and 5000 for testing. We follow the experimental settings in [11]. The results are shown in Table 3. As we can see that GFR+EBSF can have a lower detection error rate compared to SRNet. And the improvement is greater at low payloads than at high payloads. This means our proposed manual feature-based methods can achieve comparable or better with respect to deep learning based steganalysis.

5. CONCLUSION

In this paper, we propose a framework using feature combination that detects robust steganography based on repetitive JPEG compression. Through analysis, we find the correlation between steganalysis feature and compression feature and combine the two together. In experiments, we can not only achieve better results than the manual steganalysis feature but also better results than deep learning-based steganalysis.

In the future, there are still some issues that need to be investigated. Not only traditional feature sets, such as DCTR and GFR, but also deep learning-based steganalysis can be combined with different compressed features. In addition, how to perform better feature fusion is also a problem that needs to be investigated in the future.

6. REFERENCES

[1] Tomáš Filler, Jan Judas, and Jessica Fridrich, "Minimizing additive distortion in steganography using syndrome-trellis codes," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 3, pp. 920–935, 2011.

[2] Jessica Fridrich, *Steganography in digital media: principles, algorithms, and applications*, Cambridge University Press, 2009.

[3] Yi Zhang, Xiangyang Luo, Chunfang Yang, Dengpan Ye, and Fenlin Liu, "A framework of adaptive steganography resisting jpeg compression and detection," *Security and Communication Networks*, vol. 9, no. 15, pp. 2957–2971, 2016.

[4] Yi Zhang, Xiaodong Zhu, Chuan Qin, Chunfang Yang, and Xiangyang Luo, "Dither modulation based adaptive steganography resisting JPEG compression and statistic detection," *Multimedia Tools and Applications*, vol. 77, no. 14, pp. 17913–17935, 2018.

[5] Xinzhi Yu, Kejiang Chen, Yaofei Wang, Weixiang Li, Weiming Zhang, and Nenghai Yu, "Robust adaptive steganography based on generalized dither modulation and expanded embedding domain," *Signal Processing*, vol. 168, pp. 107343, 2020.

[6] Zengzhen Zhao, Qingxiao Guan, Hong Zhang, and Xianfeng Zhao, "Improving the robustness of adaptive steganographic algorithms based on transport channel matching," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 7, pp. 1843–1856, 2018.

[7] Vojtěch Holub and Jessica Fridrich, "Low-complexity features for JPEG steganalysis using undecimated dct," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 2, pp. 219–228, 2014.

[8] Xiaofeng Song, Fenlin Liu, Chunfang Yang, Xiangyang Luo, and Yi Zhang, "Steganalysis of adaptive JPEG steganography using 2d gabor filters," in *Proceedings of the 3rd ACM workshop on information hiding and multimedia security*, 2015, pp. 15–23.

[9] Vojtěch Holub and Jessica Fridrich, "Phase-aware projection model for steganalysis of JPEG images," in *Media Watermarking, Security, and Forensics 2015*. International Society for Optics and Photonics, 2015, vol. 9409, p. 94090T.

[10] Mehdi Yedroudj, Frédéric Comby, and Marc Chaumont, "Yedroudj-net: An efficient cnn for spatial steganalysis," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 2092–2096.

[11] Mehdi Boroumand, Mo Chen, and Jessica Fridrich, "Deep residual network for steganalysis of digital images," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 5, pp. 1181–1193, 2018.

[12] Yiwei Zhang, Weiming Zhang, Kejiang Chen, Jiayang Liu, Yujia Liu, and Nenghai Yu, "Adversarial examples against deep neural network based steganalysis," in *Proceedings of the 6th ACM Workshop on information hiding and multimedia security*, 2018, pp. 67–72.

[13] Zhaoxia Yin and Longfei Ke, "Robust adaptive steganography based on dither modulation and modifcation with re-compression," *IEEE Transactions on Signal and Information Processing over Networks*, 2021.

[14] Fengyong Li, Kui Wu, Chuan Qin, and Jingsheng Lei, "Anti-compression jpeg steganography over repetitive compression networks," *Signal Processing*, vol. 170, pp. 107454, 2020.

[15] Jianquan Yang, Jin Xie, Guopu Zhu, Sam Kwong, and Yun-Qing Shi, "An effective method for detecting double JPEG compression with the same quantization matrix," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 11, pp. 1933–1942, 2014.

[16] Fangjun Huang, Jiwu Huang, and Yun Qing Shi, "Detecting double jpeg compression with the same quantization matrix," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 4, pp. 848–856, 2010.

[17] Vojtěch Holub and Jessica Fridrich, "Digital image steganography using universal distortion," in *Proceedings of the first ACM workshop on Information hiding and multimedia security*, 2013, pp. 59–68.

[18] Jan Kodovsky, Jessica Fridrich, and Vojtěch Holub, "Ensemble classifiers for steganalysis of digital media," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 2, pp. 432–444, 2011.

[19] Patrick Bas, Tomáš Filler, and Tomáš Pevnỳ, "" break our steganographic system": the ins and outs of organizing boss," in *International workshop on information hiding*. Springer, 2011, pp. 59–70.

[20] P Bas and T Furon, "Bows-2 contest (break our watermarking system)," *organised within the activity of the Watermarking Virtual Laboratory (Wavila) of the European Network of Excellence ECRYPT*, 2008.

[21] Rémi Cogranne, Vahid Sedighi, Jessica Fridrich, and Tomáš Pevnỳ, "Is ensemble classifier needed for steganalysis in high-dimensional feature spaces?," in *2015 IEEE International Workshop on Information Forensics and Security (WIFS)*. IEEE, 2015, pp. 1–6.