JPEG Image Steganalysis Using Weight Allocation from Block Evaluation*

Weiwei LUO†, Wenpeng ZHOU††, Jinglong FANG††††, Nonmembers, and Lingyan FAN††††, Member

SUMMARY. Recently, channel-aware steganography has been presented for high security. The corresponding selection-channel-aware (SCA) detecting algorithms have also been proposed for improving the detection performance. In this paper, we propose a novel detecting algorithm of JPEG steganography, where the embedding probability and block evaluation are integrated into the new probability. This probability can embody the change due to data embedding. We choose the same high-pass filters as maximum diversity cascade filter residual (MD-CFR) to obtain different image residuals and a weighted histogram method is used to extract detection features. Experimental results on detecting two typical steganographic algorithms show that the proposed method can improve the performance compared with the state-of-art methods.

key words: JPEG image, steganalysis, block evaluation, weight allocation

1. Introduction

JPEG image is widely used in the current social network and many data hiding methods are designed over the JPEG image. To detect whether the extra data are embedded in the JPEG image, JPEG steganalysis uses the machine learning technique for detecting this behavior. With the complexity of steganographic algorithms increasing, image steganalysis needs to start designing the effective high-dimensional feature. To improve the detection performance of the steganographic method, Kodovsky et al. provide many models from the joint distribution of DCT coefficients to generate 11255-dimensional features. After Cartesian calibration [2], total double dimensions of 22510-dimensional features are named Cartesian-calibrated JPEG rich model features (CC-JRM). CC-JRM is considered as the proper selection of detection features on DCT coefficients. While obtaining the high-dimensional feature, there are many classifiers to be used. The Fisher discriminant analysis (FDA) ensemble classifier [1] is widely utilized due to low complexity and good performance. This ensemble method extracts feature subsets followed by the Bagging technique and generates the diverse results.

Recently, researchers more pay attention on spatial coefficients because the features extracted from the spatial image explore obviously the trace after embedding data. These embedding data are considered as the high-frequency signal, hence most of research employ the high-pass filter to obtain the image residual. A pioneering work firstly converts a JPEG image into the spatial image [3], and uses 64-DCT filters to extract histogram features via fusing the some symmetric positions named discrete cosine transform residual (DCTR) feature. Later, main results on JPEG steganalysis follow this rule and devise the better filters to obtain the effective residual which can embody the difference before and after steganography. A typical work is based on multiple directional Gabor filters and extracts 17000-dimensional Gabor filter residual (GFR) features [4]. Its detection performance is better than the DCTR feature and is often used to evaluate steganographic algorithms.

Nowadays, two cascade filter sets which have diverse attribute are proposed in maximum diversity cascade filter residual (MD-CFR) [6]. The direction of filters mainly depend on basic filters in the filter base and the scale of filters is expressed through the autoconvolution. This method gives a new idea that people extract the detection features after defining the proper filter base. Because the JPEG image is compressed based on the block with the size of $8 \times 8$ and steganographic method embeds data into blocks one by one, the texture block is insensitive to statistical features and is used to embed the more data than the smooth block. The MD-CFR feature treats all blocks equally to extract histogram features, therefore, the complexity of the block needs to be further explored. In this paper, we fully research the block attributes and design the metric to evaluate the texture complexity of each block, moreover give the fused probability from block evaluation.

2. Related Work

Because the change from steganography can be considered as the noise on the cover image, currently popular features are exacted based on the image residual which can be obtained after filtering image using the high-pass filters. The key of detection performance is how to devise these high-pass filters. Two typical features named GFR and MD-CFR both present the different diverse filters. Here, we briefly introduce the principle of recent MD-CFR features, since
MD-CFR can achieve the competing performance.

### 2.1 High-Pass Filters in MD-CFR

As stated as before, the good high-pass filters can effectively explore the trace after embedding data. To obtain the rich residual modes, the practical features are extracted from many high-pass filters. MD-CFR suggests a framework of generating high-pass filters. These filters can capture the embedding data from the different directions and scales. In MD-CFR, authors define the two groups of filters, which are all high-pass filters. Each group of filters can be created by auto-convoluting on the basic filter set. Therefore, they only set the explicit elements in the basic set. Each base only includes eight basic filters. As an example, we list a group of basic filters in Fig. 1. Another group of filters have the same size and the different coefficients.

### 2.2 Feature Exaction and Fusion

For each of the cascade filters, the absolute values of the elements are symmetrically distributed around the center. The distributions of the corresponding residuals will be approximately symmetrical. Therefore, features extracted from the residuals with symmetrical values can be merged to reduce the dimensionality. After uncompressed a JPEG image into the spatial image named the dimensionality. After uncompressed a JPEG image into approximately symmetrical. Therefore, features extracted from the embedding data from the different directions and scales. In MD-CFR, authors define the two groups of filters, which are all high-pass filters. Each group of filters can be created by auto-convoluting on the basic filter set. Therefore, they only set the explicit elements in the basic set. Each base only includes eight basic filters. As an example, we list a group of basic filters in Fig. 1. Another group of filters have the same size and the different coefficients.

### 3. Weighted Histogram Features

In this section, we propose the weighted histogram features which depends on the texture complexity of each block. Because the modern steganography usually follows distortion minimization framework, the texture block is more embedded information than the smooth block. We need to pay attention on these texture block. Firstly we introduce the main principle of the modern steganography. The user embeds the secret data using a given distortion function. We set the embedding probability of the ith non-zero quantized DCT coefficient $c(i)$ as $p(i)$. The totally embedding capacity $L$ (bits) can be calculated by

$$L = - \sum_i \{ p(i) \log_2 p(i) + [1 - p(i)] \log_2 [1 - p(i)] \} \tag{3}$$

In the practical application, the theoretical result can be realized by the coding method [9]. As an example, the well-known JPEG domain universal wavelet relative distortion (J-UNIWARD) [10] uses a sum of relative changes of wavelet coefficients to define the embedding costs. This probability can also be used in selection-channel-aware (SCA) steganalysis [5]. As a result, the embedding data are collected into the texture block. Motivated by this idea, we design the weighted features which concentrate the texture block.

We choose the same high-pass filters as MD-CFR and obtain the image residuals. Moreover, we divide the image $I$ into the non-overlap $8 \times 8$ blocks $B_m(k, l), (m = 1, 2, \ldots, M)$, where $B_m(k, l)$ is consistent to the JPEG format on the block size. For the $m$th block $B_m(k, l)$ with the size of $8 \times 8$ indexed

$$A_m = \frac{1}{64} \sum_{k=1}^{8} \sum_{l=1}^{8} B_m(k, l) \tag{4}$$

We formulate the block texture as

$$T_m = \sqrt{\frac{1}{64} \sum_{k=1}^{8} \sum_{l=1}^{8} (B_m(k, l) - A_m)^2} \tag{5}$$

where $T_m$ is used to described the block complexity. Because $T_m$ in Eq. (5) may take a larger number, we need to control $T_m$ into a range of $(0.5, 1)$, which keeps the similar range of the probability.

$$w_m = \frac{1}{1 + e^{-\alpha T_m}} \tag{6}$$

For the $m$th block, we obtain the corresponding weighed value

$$p_m(k, l) = w_m p_m(k, l) \tag{7}$$
where \( p_m(k,l) \) is the embedding probability using the corresponding steganographic method. The embedding probabilities are obtained in the transformed domain and the feature extraction is achieved in the spatial domain. Hence, the probabilities in the DCT coefficients will be mapped the weight value in the spatial domain [5]. All \( p_m(k,l) \) can constitute the final probability image \( \hat{p}(x,y) \). In this way, a key aspect in each histogram feature is that of achieving such weighted information. We set same the fusion method with MD-CFR corresponding to the image residual.

\[
    h(r) = \frac{1}{|I|} \sum_{x,y} [\hat{p}(x,y)QT(l'(x,y))] = r
\]

where \( QT \) is same definition as Eq. (1). At this moment, we can obtain the same dimensionality with the original MD-CFR (i.e. without feature reduction). The whole processing is summarized in Algorithm 1.

\[
    \text{Algorithm 1 Feature extraction}
    \begin{align*}
        &1: \text{for } i = 1 \text{ to } M \text{ do} \\
        &2: \quad \text{Obtain the residual image through filtering the image;} \\
        &3: \quad \text{Calculate the weight of each } 8 \times 8 \text{ block;} \\
        &4: \quad \text{Select subimage and obtain the weighted histogram;} \\
        &5: \quad \text{Merge histogram and achieve the feature based on the certain residual;} \\
        &6: \quad \text{end for} \\
        &7: \text{Fuse } M \text{ residual results into high-dimension features.}
    \end{align*}
\]

4. Experimental Results

In this section, the universal image set BOSSBASE 1.01 is used to verify the proposed method. Total 10000 gray images with the size of 512 \( \times \) 512 are included. 10000 stego images are obtained through a certain steganographic algorithm with the specified payload. Random half of cover and stego images is selected as the training image set. Two quality factors of 75 and 95 in the JPEG standard are used to compress all images. To verify detecting performance, we adopt two popular JPEG steganographic algorithms, including uniform embedding revisited distortion (UERD) [8] and J-UNIWARD [10]. We extract MD-CFR [6], method in [7] and the proposed weighted features as comparable high-dimensional features for all the experiments. Each result is shown based on the average of ten trials. The entire platform of image steganalysis has been classified with the original FLD-EC [1] and the error rate \( PE \) is obtained by the average of the false alarms and missed detections probabilities.

The proposed algorithm is firstly compared with the different numbers in the feature set and different number of classifiers to affirm an optimal number. Figures 2-3 show average \( PE \) using J-UNIWARD and UERD steganography at 0.1 bpac (bits per nonzero AC DCT coefficient) with the incremental features and classifiers. In Fig. 2, we fix the number of classifiers to 129 and add extracted features, where X-axis means the number of the selected features. It is found that the performance is good when the number of features is greater than 2200, therefore we choose 2200 features in each classifier. In Fig. 3, we fix the number of extracted features to 2200 and add classifiers, where X-axis means the number of the selected classifiers. It is found that the performance is good when the number of classifiers is 139, therefore we choose 139 classifiers as fusion. Totally, we set 139 classifiers and 2200 features for all cases. For the weight \( a \) in Eq. (6), we set the different \( a \) values for searching the optimal results. We use \( PE \) for detecting J-UNIWARD steganography at 0.5 bpac payload. The results are listed in Table 1. It is found \( PE \) achieves the lowest probability when \( a = 0.8 \). Hence, we set \( a = 0.8 \) in experiments.

| Weight \( a \)  | 0.6  | 0.8  | 1.0  | 1.2  | 1.4  |
|---------------|------|------|------|------|------|
| \( PE \)      | 0.0386 | 0.0384 | 0.0387 | 0.0388 | 0.0389 |

Tables 2-3 show the average \( PE \) for detecting UERD and J-UNIWARD steganography with respect to the competing features. Sets of stego images are obtained using five distinct (0.1, 0.2, 0.3, 0.4, 0.5) bpac payloads at QF=75 and QF=95. It is observed that the proposed method usually outperforms preliminary MD-CFR, SCA-MD-CFR, and method in [7] for most variant payloads. When the payload is 0.1 bpac and 0.2 bpac, the proposed method reaches the best \( PE \) gain about 0.6% at QF=95 for J-UNIWARD steganography, but the performance gain is slight for the high payload. The main reason is that all methods are obviously detected for the high payload.
Table 2  
$P_E$ of UERD and J-UNIWARD for quality factor 75 at distinct payloads with different features.

| Features          | Payloads | 0.1     | 0.2     | 0.3     | 0.4     | 0.5     |
|-------------------|----------|---------|---------|---------|---------|---------|
| MD-CFR [6]        |          | 0.367   | 0.226   | 0.128   | 0.072   | 0.040   |
| Method in [7]     |          | 0.363   | 0.223   | 0.126   | 0.067   | 0.036   |
| SCA-MD-CFR [6]    |          | 0.245   | 0.130   | 0.076   | 0.043   | 0.025   |
| Proposed          |          | 0.244   | 0.131   | 0.074   | 0.042   | 0.023   |

| Features          | Payloads | 0.1     | 0.2     | 0.3     | 0.4     | 0.5     |
|-------------------|----------|---------|---------|---------|---------|---------|
| MD-CFR [6]        |          | 0.404   | 0.276   | 0.168   | 0.095   | 0.050   |
| Method in [7]     |          | 0.400   | 0.270   | 0.160   | 0.090   | 0.047   |
| SCA-MD-CFR [6]    |          | 0.348   | 0.218   | 0.127   | 0.071   | 0.040   |
| Proposed          |          | 0.346   | 0.216   | 0.123   | 0.071   | 0.038   |

Table 3  
$P_E$ of UERD and J-UNIWARD for quality factor 95 at distinct payloads with different features.

| Features          | Payloads | 0.1     | 0.2     | 0.3     | 0.4     | 0.5     |
|-------------------|----------|---------|---------|---------|---------|---------|
| MD-CFR [6]        |          | 0.446   | 0.364   | 0.280   | 0.199   | 0.135   |
| Method in [7]     |          | 0.444   | 0.361   | 0.278   | 0.198   | 0.132   |
| SCA-MD-CFR [6]    |          | 0.375   | 0.280   | 0.210   | 0.151   | 0.106   |
| Proposed          |          | 0.370   | 0.280   | 0.208   | 0.150   | 0.105   |

| Features          | Payloads | 0.1     | 0.2     | 0.3     | 0.4     | 0.5     |
|-------------------|----------|---------|---------|---------|---------|---------|
| MD-CFR [6]        |          | 0.473   | 0.418   | 0.338   | 0.256   | 0.180   |
| Method in [7]     |          | 0.471   | 0.412   | 0.334   | 0.252   | 0.176   |
| SCA-MD-CFR [6]    |          | 0.460   | 0.395   | 0.320   | 0.248   | 0.183   |
| Proposed          |          | 0.454   | 0.389   | 0.318   | 0.246   | 0.182   |

5. Conclusion

In this paper, we present a new weighted feature to detect the steganographic behavior. The proposed weight can be calculated by using the block texture and embedding probability. We use a weighted histogram as the extracted features based on the image residual. Because the JPEG image is compressed based on each $8 \times 8$ block, we pay attention on the coded block. It is difficult to detect the steganographic behavior in the texture block, therefore, we allocate the high weights on these blocks. Experimental results verify the proposed method is better than several recent methods.

References

[1] J. Kodovský, J. Fridrich, and V. Holub, “Ensemble classifiers for steganalysis of digital media,” IEEE Transactions on Information Forensics and Security, vol.7, no.2, pp.432–444, 2012.

[2] J. Kodovský and J. Fridrich, “Steganalysis of JPEG images using rich models,” Proceedings of SPIE, Electronic Imaging, Media Watermarking, Security, and Forensics XIV, San Francisco, CA, vol.8303, p.83030A, Jan. 23–25, 2012.

[3] V. Holub and J. Fridrich, “Low complexity features for JPEG steganalysis using undecimated DCT,” IEEE Transactions on Information Forensics and Security, vol.10, no.2, pp.219–228, 2015.

[4] X. Song, F. Liu, C. Yang, X. Luo, and Y. Zhang, “Steganalysis of adaptive JPEG steganography using 2D Gabor filters,” Proceedings of the 3rd ACM Workshop on Information Hiding and Multimedia Security, pp.15–23, 2015.

[5] T. Denemark, M. Boroumand, and J. Fridrich, “Steganalysis features for content-adaptive JPEG steganography,” IEEE Transactions on Information Forensics and Security, vol.11, no.8, pp.1736–1746, 2016.

[6] G. Feng, X. Zhang, Y. Ren, Z. Qian, and S. Li, “Diversity-based cascade filters for JPEG steganalysis,” IEEE Transactions on Circuits and Systems for Video Technology, vol.30, no.2, pp.376–386, 2020.

[7] Z. Jin, G. Feng, Y. Ren, and X. Zhang, “Feature extraction optimization of JPEG steganalysis based on residual images,” Signal Processing, vol.170, no.5, pp.107455, 2020.

[8] L. Guo, J. Ni, W. Su, C. Tang, and Y.-Q. Shi, “Using statistical image model for JPEG steganography: uniform embedding revisited,” IEEE Transactions on Information Forensics and Security, vol.10, no.12, pp.2669–2680, 2015.

[9] T. Filler, J. Judas, and J. 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.

[10] V. Holub, J. Fridrich, and T. Denemark, “Universal distortion function for steganography in an arbitrary domain,” EURASIP Journal on Information Security, vol.2014, no.1, pp.1–13, 2014.