IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 6, 2018, accepted November 26, 2018, date of publication November 30, 2018,
date of current version December 27, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2884198

On Improving Distortion Functions for JPEG Steganography

ZICHI WANG[1], ZHENXING QIAN[1], XINPENG ZHANG[1], MIN YANG[2], AND DENGPAN YE[3]

[1]Key Laboratory of Specialty Fiber Optics and Optical Access Networks, Joint International Research Laboratory of Specialty Fiber Optics and Advanced Communication, Shanghai Institute for Advanced Communication and Data Science, Shanghai University, Shanghai 200444, China
[2]School of Computer Science and Technology, Fudan University, Shanghai 201203, China
[3]School of Cyber Science and Engineering, Wuhan University, Wuhan 430072, China

Corresponding author: Zhenxing Qian (zxqian@shu.edu.cn)

This work was supported in part by the Natural Science Foundation of China under Grants U1536108, 61572308, U1636206, U1736213, UI636101, U1636219, U1736211, and 61373151, in part by the National Key Research Development Program of China under Grant 2016QY01W0200, in part by the Shanghai Excellent Academic Leader Plan under Grant 16XD1401200, and in part by the Natural Science Foundation of Shanghai under Grant 18ZR1427500.

ABSTRACT This paper proposes a framework to improve the existing distortion functions designed for JPEG steganography, which results in a better capability of countering steganalysis. Different from the existing steganography approach that minimizes image distortion, we minimize the feature distortion caused by data embedding. Given a JPEG image, we construct a reference image close to the image before JPEG compression. Guided by both the reference image and the feature distortion minimization, the state-of-the-art distortion functions designed for syndrome trellis coding embedding are improved by distinguishing the embedding costs for $+1$ versus -1 embedding. This paper has three contributions. First, the proposed framework outperforms the traditional, since we use the constructed reference image and the public steganalytic knowledge for data embedding. Second, the proposed framework is universal for improving distortion functions that were designed for JPEG steganography. Finally, experimental results also prove that the proposed approach has a better undetectability when examined by modern steganalytic tools.

INDEX TERMS Steganography, JPEG, distortion function.

I. INTRODUCTION

Steganography is a technology of hiding secret information into a digital media, aiming at transmitting the secret data through public channels without drawing suspicion [1]. On the contrary, steganalysis is a technology of disclosing secret transmission by analyzing the media on public channels [2]. Therefore, the capability of countering steganalysis is vital for steganography. In the past decades, many approaches have been developed for both technologies.

Digital images are widely used as covers in steganography. In the early techniques of digital image steganography, secret data is hidden by modifying the LSB (Least Significant Bit) of an image to keep good visual quality [3]. These methods can be broken by dedicated steganalysis according to the abnormal properties, e.g., pixel histogram, JPEG artifacts, or color palette [4], [5]. Subsequently, some improved methods are designed to keep a part of statistical property unchanged [6]–[9]. However, the security performances are still not satisfactory. For instance, even though the LSB matching [6] can avoid the abnormal properties in pixel histogram, the hidden data can be disclosed according to the other abnormalities, e.g., the variation of centroid of Fourier coefficients [10], the variation of statistical properties in the two LSB planes [11], or the noise distribution in decompressed images [12].

To achieve better performances on security, steganography coding algorithms were proposed to decrease the modifications in an image caused by data hiding [13]–[15], such as matrix embedding [13], EMD (Exploiting Modification Direction) coding [14], and ZZW (Zhang, Zhang, and Wang) coding [15]. These means are efficient to combat the dedicated steganalytic tools. However, they are not secure enough against modern steganalysis, since the features of an image are not thoroughly investigated during data embedding.

Recent works on image steganography focus on adaptive embedding thanks to the use of STC (Syndrome Trellis Coding) based embedding [16], which try to minimize the additive distortion between a cover and a stego using

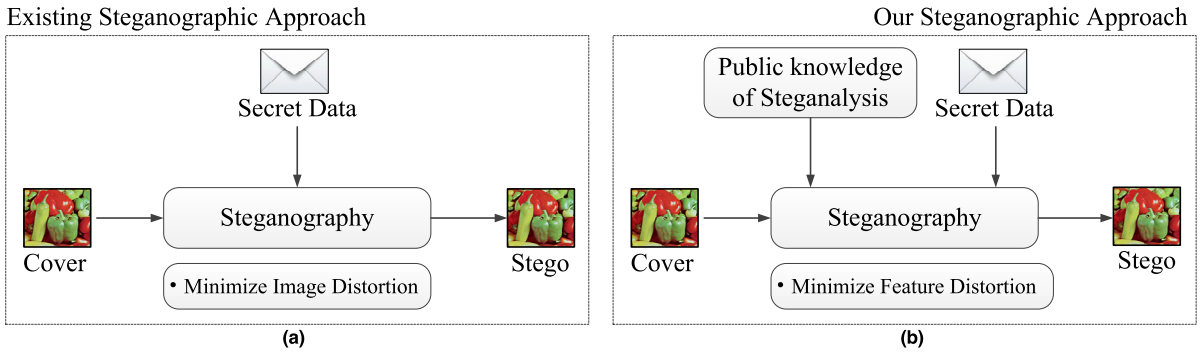

FIGURE 1. Existing and our steganographic approach.

a predefined distortion function. The distortion function assigns embedding costs for all elements in a cover, i.e., pixels for an uncompressed image, or DCT (Discrete Cosine Transform) coefficients for a JPEG image. These embedding costs are used to quantify the effects when modifying the cover elements. Many good distortion functions have been proposed for spatial images, such as WOW (Wavelet Obtained Weights) [17], SUNIWARD (Spatial UNIversal WAvelet Relative Distortion) [18], [19], HILL (High-pass, Low-pass, and Low-pass) [20], MiPOD (Minimizing the Power of Optimal Detector) [21], and ASO (Adaptive Steganography by Oracle) [22]. Since JPEG is popular for digital images, many distortion functions have also been developed by JPEG steganography, e.g., JUNIWARD JPEG UNIversal WAvelet Relative Distortion) [18], [19], UED (Uniform Embedding Distortion) [23], UERD (Uniform Embedding Revisited Distortion) [24], IUERD (Improved UERD) [25], HDS (Hybrid Distortion Steganography) [26], and RBV (Residual Blocks Value) [27].

On the other hand, with the development of machine learning, many steganalytic tools were proposed to detect the behaviors of steganography. Most steganalytic methods use supervised machine learning to investigate the models of the covers and the stegos. The adversary extracts features from a set of images to train a common steganalytic model, which is then used to distinguish the suspicious images [28]–[31]. Recently, deep learning based steganalysis achieved good performances [32]–[34]. However, the study of deep-learned features is still in its infancy [35]. Some important operations in the feature extraction process, e.g., truncation and quantization, cannot be effectively learned by deep networks. Hence, the feature extraction is still important for steganalysis. Many popular feature sets have been proposed for JPEG steganalysis, such as ccJRM (Cartesian Calibrated JPEG Rich Model) [36], DCTR (Discrete Cosine Transform Residual) [37], GFR (Gabor Filters Residual) [38], and the improved GFR [39]. During classification, the ensemble classifier is popular for measuring the feature sets [40], [41]. When analyzing state-of-the-art steganographic methods, modern steganalytic approaches are much more powerful than traditional tools.

Nowadays, most steganographic methods aim at defining better distortion functions to achieve better capabilities of defeating modern steganalysis. However, the public knowledge of steganalysis (the steganalysis tools or technologies which can be handled) has not been involved during STC based embedding. As shown in Fig. 1(a), the most STC based steganography tries to minimize steganographic distortion on a cover without any knowledge of steganalysis. In this paper, we discuss another steganography approach that the public knowledge of steganalysis is used by a steganographer [42]. The public knowledge of steganalysis means the steganalytic tools that have been published and are known to everyone. As shown in Fig. 1(b), we design the steganographic method with the guide of steganalysis. Different from the existing steganography by minimizing image distortions, we minimize the distortion between the cover features and the stego features. This way, it would be more difficult for a steganalyzer to distinguish the stego from the cover. Therefore, the undetectability of steganography can be greatly improved. The proposed method is an example of ‘‘adversarial signal processing’’ [1].

The rest of this paper is organized as follows. We introduce the related works in Section II. The proposed framework described in Section III. Experimental results and analysis are provided in Section IV. Section V concludes the whole paper.

II. RELATED WORK

In this Section, some related works are introduced, including distortion function design, distortion function improvement, and feature sets used in JPEG steganalysis.

A. DISTORTION FUNCTIONS FOR JPEG STEGANOGRAPHY

When using STC based data embedding, many distortion functions have been designed for JPEG steganography, e.g., JUNIWARD [18], [19], UED [23], UERD [24], and HDS [26].

The distortion function of JUNIWARD is computed as a sum of relative changes of wavelet coefficients, in which wavelet directional filter banks are used to decompose the cover image. Because of the directionality of wavelet basis functions, the embedding changes concentrates on some

regions, e.g., the texture or noisy regions. As a result, it is difficult for a steganalyzer to construct models for multiple directions. However, enormous computational complexity was required by the wavelet filter banks in JUNIWARD.

To save the computation complexity, UED and UERD were proposed to spread embedding modifications uniformly to DCT coefficients of all possible magnitudes. UED was the first attempt to define the embedding cost for DCT coefficients and implement uniform embedding strategy. The distortion function in UED is derived directly from the quantized DCT coefficients. The embedding cost for each quantized DCT coefficient is the reciprocal of a sum of the magnitude of the DCT coefficient and its intra- and inter-block neighborhood coefficients. Nevertheless, the undetectability of UED is far below JUNIWARD. After that, UERD was proposed to achieve a better result of undetectability than UED. The block distortions and quantization steps are used to refine the uniform embedding strategy. The block distortion is the reciprocal of block energy, which is the sum of absolute values of DCT coefficients. The embedding cost for each quantized DCT coefficient is the product of block distortion and quantization step.

In UERD, the block distortions were adopted and incarnated in DCT domain. Actually, the block distortion can be better measured in spatial domain using pixel correlations. To this end, a hybrid distortion function for JPEG image exploiting block fluctuation and quantization steps was proposed in [26]. According to the block fluctuation, a distortion value is assigned for each 8×8 block in the spatial domain. Meanwhile, quantization steps are used to assign distortion values for DCT coefficients in each block. Both measurements are combined together via multiplication to evaluate the effects when modifying the cover elements.

B. IMPROVEMENT OF DISTORTION FUNCTION

Besides the definition of distortion functions, some schemes have been proposed to improve the JPEG distortion functions with the help of additional information of the precover [43], [44] or the same scene based multiple JPEG images [45].

In [43], a principle of incorporating the side information with STC based steganography was proposed. For a JPEG image, the unquantized DCT coefficients generated by compressing the precover are used as side information to improve the distortion function. The embedding costs for changing a DCT coefficient X_{ij} into Y_{ij} by $+1$ and -1 are shown in (1) and (2), respectively,

$$\rho_{ij}^{(SI)+} = (1 - 2\left|e_{ij}\right|)\rho_{ij}^{(\mathrm{A})} \quad \text{if } Y_{ij} = X_{ij} + \mathrm{sign}(e_{ij}) \tag{1}$$

$$\rho_{ij}^{(SI)-} = \rho_{ij}^{(\mathrm{A})} \quad \text{if } Y_{ij} = X_{ij} - \mathrm{sign}(e_{ij}) \tag{2}$$

where e_{ij} is the quantization error, $\rho_{ij}^{(A)}$ is the cost defined by existing works, and $\mathrm{sign}(\cdot)$ is the sign function. The improved distortion function represents both the local image complexity and the quantization distortion. Once the distortion is decreased, there would be fewer changes of statistical properties in the covers after data embedding.

Recently, Denemark and Fridrich [45] propose an alternative steganography based on side information, in which multiple JPEG images sharing the same scene are used as references. Those reference JPEG images are used to determine the polarities of embedding changes when modifying DCT coefficients. Let the DCT coefficients in a cover and a reference image be $x_{ij}^{(1)}$ and $x_{ij}^{(2)}$, respectively. For the coefficients that satisfy $x_{ij}^{(1)} = x_{ij}^{(2)}$, the preliminarily defined costs are kept unchanged. Otherwise, the costs should be improved by modulation. Let the original cost be $\rho_{ij}^{(0)}(\pm 1)$. The improved cost $\rho_{ij}(\pm 1)$ can be calculated by two steps in (3) and (4),

$$\text{Step 1: set } \rho_{ij}(\pm 1) = \rho_{ij}^{(0)}(\pm 1) \tag{3}$$

$$\text{Step 2: } x_{ij}^{(1)} \neq x_{ij}^{(2)} \Rightarrow \rho_{ij}(s_{ij}) = m(Q)\rho_{ij}^{(0)}(s_{ij}) \tag{4}$$

where $s_{ij} = \mathrm{sign}(x_{ij}^{(2)} - x_{ij}^{(1)})$, and $m(Q)$ is a modulation factor corresponding to the JPEG quality factor.

C. MODERN JPEG STEGANALYSIS

On the adversary side, many steganalytic methods have been proposed to defeat steganography. Feature extraction and machine learning based steganalysis has been proved to be efficient. A variety of feature extraction algorithms have been proposed for JPEG steganalysis, such as ccJRM, DCTR, and GFR. Meanwhile, the ensemble classifier is widely used to measure the feature sets.

In ccJRM, many co-occurrence matrices are calculated to describe the dependencies among different DCT coefficients. The absolute values of DCT coefficients in JPEG are treated as 64 parallel channels with weak dependencies. On the one hand, the 64 channels are modeled using co-occurrence matrices separately and collected together. On the other hand, to find the integral joint statistics among coefficients, co-occurrence matrices constructed from the whole DCT plane with a wider range of values are also collected. Meanwhile, the joint statistics are symmetrized to construct a compact and robust model. Besides, both parts of features are further diversified by modeling the differences calculated in different directions.

Different from co-occurrence matric based ccJRM, DCTR extracts features from the quantized noise in a JPEG image. After decompressing a JPEG image, a steganalyzer uses 64 convolution kernels to calculate the filtering residuals and construct histograms. Symmetries are also used to decrease the feature dimensionality. The use of histograms leads to lower dimensionality in comparison with ccJRM, and thus performs lower computational complexity.

GFR is a steganalysis algorithm based on two dimensional Gabor filters. A decompressed JPEG image is first filtered by Gabor filters with different scales and orientations. Then, the histogram features are extracted from all image filtering coefficients. Different from DCTR that uses 64 DCT kernels for image filtering, the 2D Gabor filters can capture the

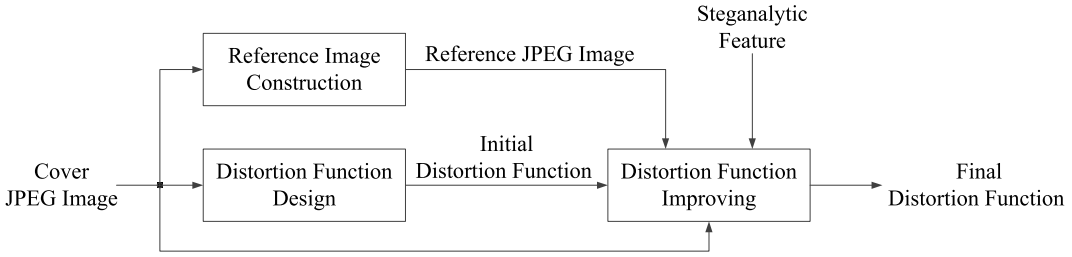

FIGURE 2. Proposed framework.

embedding changes from more scales and orientations, since the 2D Gabor filter is a local band-pass filter with optimal joint localization properties. Therefore, the feature of image filtering coefficients by GFR is more effective for detecting the adaptive JPEG steganographic methods.

III. PROPOSED FRAMEWORK OF JPEG STEGANOGRAPHY

The proposed framework is shown in Fig. 2. Given a JPEG image, we first generate an initial distortion function using a STC based steganography for JPEG. Meanwhile, a reference image which is similar to the corresponding uncompressed version is constructed. Guided by both the reference image and the public knowledge of steganalytic tools, we propose an algorithm to improve the initial distortion function.

A. REFERENCE IMAGE CONSTRUCTION

Inspired by the pioneering work in [45], which uses multiple images sharing the same scene as references, we also use reference image to identify the direction of distortion function improving. Considering there is only one image available, the reference image should be constructed instead of employ directly. To achieve satisfactory performance, the constructed reference image should be similar to the original uncompressed image. The JPEG quantization noise is non-additive. It is a problem to remove non-additive noise. However, it has been proved that the JPEG noise can be approximately modeled as Gaussian-like distribution [46], which is suitable to Wiener filter. Wiener filter is simple enough that achieves minimal MSE (mean squared error). Furthermore, as the empirical results indicated in [47], Wiener filter is a good tool for JPEG filtering. Therefore, we use this filter to generate a reference image closing to the original. Once the proper parameters are used, the reference image is close to the original image before JPEG compression.

We denote the uncompressed image as $\mathbf{I}$ that contains $M \times N$ pixels. After JPEG compression, a new image $\tilde{\mathbf{I}}$ is generated. The procedure of JPEG compression and decompression can be regarded as a noisy channel. We name this noise as JPEG noise, which can be calculated by $\mathbf{I} - \tilde{\mathbf{I}}$. Let the reference image be $\hat{\mathbf{I}}$. The MSE between $\mathbf{I}$ and $\hat{\mathbf{I}}$ is,

$$MSE = \mathrm{E}[I(n) - \hat{I}(n)]^2 \tag{5}$$

where $\mathrm{E}[\cdot]$ is the operator of mathematical expectation, and $n = 1, 2, \ldots, MN$.

To realize a Wiener filter, we can estimate the mean and standard deviation using a specified local neighborhood of each pixel. In detail, the current image is filtered using pixel-wise adaptive Wiener filtering. As described in (6) and (7), the mean μ and standard deviation σ are calculated using neighborhoods with a given size φ which will be discussed later.

$$\mu = \frac{1}{\varphi^2} \sum_{i,j \in \eta} b_{i,j} \tag{6}$$

$$\sigma = \sqrt{\frac{1}{\varphi^2} \sum_{i,j \in \eta} b_{i,j}^2 - \mu^2} \tag{7}$$

where $b_{i,j}$ is the (i, j)th pixel of image $\tilde{\mathbf{I}}$, and η is the $\varphi \times \varphi$ neighborhood of $b_{i,j}$. Meanwhile, the variance of JPEG noise, which is needed in Wiener filter, can be estimated using σ^2 calculated from the current image. Finally, we can construct a reference image $\hat{\mathbf{I}}$ using (8),

$$\hat{\mathbf{I}} = \tilde{\mathbf{I}} \otimes \mathbf{H}_0 \tag{8}$$

where $\mathbf{H}_0$ is a 2D Wiener filter. The filtered version $\hat{\mathbf{I}}$ is very close to the original image $\mathbf{I}$.

To find empirical sizes for the filters, we conduct a group of experiments over 1000 grayscale images arbitrarily selected from BOSSbase ver. 1.01 [48]. These images are compressed into JPEG with the QF (quality factor) of 55, 65, 75, 85, and 95, respectively. These generated JPEG images are further decompressed into spatial images. After Wiener filtering, we obtain 1000 reference images for each QF case. The average of PSNR values of the 1000 reference image are listed in Table 1.

The results in Table 1 show that the 3×3 filter corresponding to the best quality of reference images. Generally, the pixels of eight neighborhoods are the most correlated in a natural image, a 3×3 filter is appropriate for removing JPEG noises. The smaller sized filter cannot take full use of the pixel correlation, and the larger sized filters may introduce misleading information from the farther pixels. The estimation error of the filtered image is inevitable since the original image before JPEG is not available. The filtered image is used as reference to guide the improving of distortion function. In this case, the quality of the filtered image is significant to the performance of the proposed scheme.

TABLE 1. Average PSNR (dB) *vs* wiener filter size.

Filter Size	2×2	3×3	4×4	5×5	6×6	7×7
QF=55	35.340	**35.931**	34.456	33.923	32.998	32.498
QF=65	35.925	**36.484**	34.811	34.217	33.222	32.708
QF=75	36.593	**37.062**	35.176	34.520	33.496	32.964
QF=85	37.433	**37.778**	35.586	34.860	33.753	33.191
QF=95	38.344	**38.475**	35.956	35.166	33.986	33.388

Therefore, we choose 3×3 as the size of Wiener filter since it achieves the best quality of filtered image.

TABLE 2. PSNR (dB) comparisons of several estimating filters.

Filter Size	Wiener filter	Average filter	Median filter	Deblurring filter
QF=55	**35.931**	33.004	33.736	29.431
QF=65	**36.484**	33.114	33.942	29.477
QF=75	**37.062**	33.205	34.131	29.531
QF=85	**37.778**	33.279	34.331	29.572
QF=95	**38.475**	33.329	34.527	29.604

For other cover image estimating filters, e.g., Average filter, Median filter, and Deblurring filter, the average PSNR values over 1000 arbitrarily selected images are shown in Table 2. The size of the other three filters is also 3×3. It shows that these estimating filters are inferior than Wiener filter for all cases. Thus, Wiener filter is quite suitable for JPEG filtering in the sense of MSE minimization.

B. INEQUIVALENCE OF BIPOLAR EMBEDDING BASED MODIFICATION

Traditionally, most distortion functions for stegano- graphy share the same embedding costs for both $+1$ and -1 operations. We discovered empirically in [49] that the embedding costs for $+1$ and -1 should not be equivalent. However, no theoretical analysis has been provided directly. In this section, we provided the reasons why it works.

During the procedure of JPEG compression, an image $\mathbf{I}$ is transformed into DCT coefficients $x_{i,j}$, which is further quantized into the coefficients $c_{i,j}$, where $i = 1, 2, \ldots, M$ and $j = 1, 2, \ldots, N$. The results are represented as (9) and (10),

$$x_{i,j} = q_{i,j} \cdot \left\lfloor \frac{x_{i,j}}{q_{i,j}} \right\rfloor + \mathrm{mod}(x_{i,j}, q_{i,j}) \tag{9}$$

$$c_{i,j} = \mathrm{round}\left(\frac{x_{i,j}}{q_{i,j}}\right) \tag{10}$$

where $q_{i,j}$ is the quantization step for the (i, j)-th coefficient, and the operators $\lfloor \cdot \rfloor$ and round$(\cdot)$ stand for the floor and rounding operations, respectively. During the procedure of decompression, the coefficient $c_{i,j}$ is de-quantized as,

$$\begin{aligned}\hat{x}_{i,j} &= q_{i,j} \cdot c_{i,j} \\ &= q_{i,j} \cdot \left\lfloor \frac{x_{i,j}}{q_{i,j}} \right\rfloor + q_{i,j} \cdot \mathrm{round}\left(\frac{\mathrm{mod}(x_{i,j}, q_{i,j})}{q_{i,j}}\right)\end{aligned} \tag{11}$$

Thus, the quantization error $e_{i,j}$ is equal to,

$$\begin{aligned}e_{i,j} &= \left|x_{i,j} - \hat{x}_{i,j}\right| \\ &= \left|\mathrm{mod}(x_{i,j}, q_{i,j}) - q_{i,j} \cdot \mathrm{round}\left(\frac{\mathrm{mod}(x_{i,j}, q_{i,j})}{q_{i,j}}\right)\right| \le \frac{q_{i,j}}{2}\end{aligned} \tag{12}$$

During data hiding, the coefficient $c_{i,j}$ is modified into $c^+_{i,j}$ or $c^-_{i,j}$, representing the $+1$ or -1 modifications, respectively. Consequently, the quantization errors $e^+_{i,j}$ and $e^-_{i,j}$ are,

$$e^+_{i,j} = \left|\mathrm{mod}(x_{i,j}, q_{i,j}) - q_{i,j} \cdot \mathrm{round}\left(\frac{\mathrm{mod}(x_{i,j}, q_{i,j})}{q_{i,j}}\right) - q_{i,j}\right| \tag{13}$$

$$e^-_{i,j} = \left|\mathrm{mod}(x_{i,j}, q_{i,j}) - q_{i,j} \cdot \mathrm{round}\left(\frac{\mathrm{mod}(x_{i,j}, q_{i,j})}{q_{i,j}}\right) + q_{i,j}\right| \tag{14}$$

Therefore, if $\mathrm{mod}(x_{i,j}, q_{i,j}) \le q_{i,j}/2$, the set of inequalities in (15) is satisfied. Otherwise, the set (16) is satisfied.

$$\left\{e_{i,j} \le \frac{q_{i,j}}{2} \le e^+_{i,j} \le q_{i,j}, q_{i,j} \le e^-_{i,j} \le \frac{3q_{i,j}}{2}\right\} \tag{15}$$

$$\left\{e_{i,j} \le \frac{q_{i,j}}{2} \le e^-_{i,j} \le q_{i,j}, q_{i,j} \le e^+_{i,j} \le \frac{3q_{i,j}}{2}\right\} \tag{16}$$

The results show that the increments of quantization errors for $+1$ and -1 embedding are different, which indicates that the costs for $+1$ and -1 embedding should be defined according to the values of the original coefficients.

In real applications, as the original image before JPEG is not available, we construct a reference image that is close to the original. With this image, we improve existing distortion functions. As the aforementioned algorithm in *A* of this section, we construct the reference image by Wiener filtering. As the reference image is close the original image before JPEG compression, we use the DCT coefficients of the reference image as the evidence of modification directions. Meanwhile, we use public knowledge of steganalysis to decide the extent of modifications. Next, we detail the algorithm of distortion function improvement.

C. DISTORTION FUNCTION IMPROVEMENT

Let $\rho^+_{i,j}$ and $\rho^-_{i,j}$ be the embedding costs for $+1$ and -1, respectively, where $i = 1, 2, \ldots, M$ and $j = 1, 2, \ldots, N$. In traditional JPEG steganographic methods based on STC, $\rho^+_{i,j}$ is identical to $\rho^-_{i,j}$, i.e., $\rho^+_{i,j} = \rho^-_{i,j} = \rho_{i,j}$. We improve the costs according to the DCT coefficients $\hat{c}_{i,j}$ in the reference image $\hat{\mathbf{I}}$. As $\hat{c}_{i,j}$ is unknown to the steganalyzer, this side information is an advantage of the steganographer. As mentioned

in [45], $\hat{c}_{i,j}$ partially compensates for the lack of knowledge of the cover model when it is highly non-stationary.

With state-of-the-art algorithms, we first generate an initial embedding cost $\rho_{i,j}$ for each DCT coefficient using existing distortion functions such as JUNIWARD, UED, or UERD. Since the reference image is close to the original uncompressed image, it is advisable to modify $c_{i,j}$ toward $\hat{c}_{i,j}$ if any modification is needed during embedding. For this reason, cost $\rho_{i,j}$ is then improved by (17) and (18),

$$\rho_{i,j}^{+} = \alpha^{(1+\beta_{i,j})/2} \cdot \rho_{i,j} \tag{17}$$

$$\rho_{i,j}^{-} = \alpha^{(1-\beta_{i,j})/2} \cdot \rho_{i,j} \tag{18}$$

where α is a parameter for adjusting the extent of modification, $0 \leq \alpha \leq 1$, and $\beta_{i,j} = \text{sign}(\hat{c}_{i,j} - c_{i,j})$. A large α means $\rho_{i,j}$ is modified by a small amplitude, and vice versa.

We define the $+1$ embedding cost $\rho_{i,j}^{+}$ by decreasing the initial embedding cost $\rho_{i,j}$ when $c_{i,j} < \hat{c}_{i,j}$. Otherwise, we define $\rho_{i,j}^{-}$ by decreasing the initial cost when $c_{i,j} > \hat{c}_{i,j}$. Generally, $c_{i,j}$ is not identical to $\hat{c}_{i,j}$ since $c_{i,j}$ is an integer value while $\hat{c}_{i,j}$ is a real one.

Next, we determine the value of the parameter α using the public knowledge of steganalysis. We denote the feature vectors of cover and a stego image as $\mathbf{F^c} = \{f_r^c\}$ and $\mathbf{F}^s = \{f_r^s\}$, respectively, where $r \in \{1, 2, \ldots, w\}$, and w represents the feature dimension. The Euclidean distance between $\mathbf{F^c}$ and $\mathbf{F}^s$ can be calculated by

$$d = \sqrt{\sum_{r=1}^{w} (f_r^c - f_r^s)^2} \tag{19}$$

To identify a parameter α such that the Euclidean distance is minimized, we propose to use the following algorithm. The factor α is determined by the cover image itself. We assign different α values for different images by choosing the best one corresponding to the shortest feature distance. The security is related to the modification amplitude. However, a strong or a small modification does not guarantee an increment of the undetectability. Only a suitable modification can result in a shortest feature distance.

FIGURE 3. Test image.

With different steganalysis knowledge, the values of α are different. We conduct a group of experiments on the figure shown in Fig. 3, which is arbitrarily selected from BOSSbase ver. 1.01 [48] and compressed into JPEG using a quality factor QF $= 75$. We use UED, UERD, and JUNIWARD to embed the payloads of 0.1 and 0.4 bpnzac (bit per non-zero AC coefficient). The determined values of parameter α corresponding to steganalytic feature of ccJRM, DCTR, and GFR are shown in Table 3. The results show that the parameter is determined by the used knowledge of steganalytic tools. The universal α will be given in next Section. The feature space itself provides a alternative model of covers useful for benchmark [42]. Modern high-dimensional steganalytic features employed in our method are able to complete the feature space. In this paper, reducing the distance of some specified features is empirically efficient for state-of-the-art steganalysis methods.

Algorithm 1 Identifying α

1) Set ten initial values of α as $\{0.1, 0.2, \ldots, 1\}$.
2) Based on STC steganography framework, we embed data into a given cover JPEG image using the distortion function improved by ten initial values of α, respectively. (The improved function is identical to the initial if $\alpha = 1$). Calculate ten Euclidean distances of feature vector between cover and the transitional stego images.
3) Find an α that results in a minimal distance, and denote it as $\tilde{\alpha}$.
4) To find a better parameter, we set the updated values of α as $\{\tilde{\alpha} - 0.05, \ldots, \tilde{\alpha} - 0.01, \tilde{\alpha} + 0.01, \ldots, \tilde{\alpha} + 0.05\}$.
5) Embed data into the cover image with the distortion function improved by the ten updated values of α, respectively. Calculate ten Euclidean distances of feature vectors between cover and the transitional stego images.
6) Find an α among the eleven updated values (Including $\tilde{\alpha}$) that corresponds to a minimal distance.

TABLE 3. Parameter α corresponding to steganalytic features.

	ccJRM	DCTR	GFR
UED (0.1 bpnzac)	0.76	0.92	0.71
UED (0.4 bpnzac)	0.93	0.82	0.64
UERD (0.1 bpnzac)	0.92	0.85	0.88
UERD (0.4 bpnzac)	0.90	0.88	0.58
JUNIWARD (0.1 bpnzac)	0.91	0.95	0.74
JUNIWARD (0.4 bpnzac)	0.68	0.65	0.63

D. DATA EMBEDDING AND EXTRACTION

After improving the distortion function, we use the ternary STC framework embed secret bits. The theoretical minimal steganography distortion D of stego image with embedding capacity γ (bits) [50] is

$$D = \sum_{i=1}^{M} \sum_{j=1}^{N} (p_{i,j}^{+} \rho_{i,j}^{+} + p_{i,j}^{-} \rho_{i,j}^{-}) \tag{20}$$

where

$$p_{i,j}^{+}=\frac{e^{-\lambda\rho_{i,j}^{+}}}{1+e^{-\lambda\rho_{i,j}^{+}}+e^{-\lambda\rho_{i,j}^{-}}} \text{ and } p_{i,j}^{-}=\frac{e^{-\lambda\rho_{i,j}^{-}}}{1+e^{-\lambda\rho_{i,j}^{+}}+e^{-\lambda\rho_{i,j}^{-}}}$$

are the probabilities of modifying $c_{i,j}$ by $+1$ or -1 with constraint $0 < p_{i,j}^{+}+p_{i,j}^{-} < 1$. The parameter λ ($\lambda > 0$) is used to make the ternary information entropy of modifying probability equal to the capacity γ,

$$-\sum_{i=1}^{M}\sum_{j=1}^{N}\Big\{p_{i,j}^{+}\log_2 p_{i,j}^{+}+p_{i,j}^{-}\log_2 p_{i,j}^{-} +(1-p_{i,j}^{+}-p_{i,j}^{-})\log_2(1-p_{i,j}^{+}-p_{i,j}^{-})\Big\}=\gamma \quad (21)$$

The ternary STC provides a practical encoding scheme to approach this theoretical bound. Once the embedding costs of ± 1 are specified, the secret data can be embedded into the cover by ternary STC. In the phase of data embedding, the quantized DCT coefficients of a cover image, the distortion function, and secret data are inputted into STC [16] (with a constraint height $h = 10$) to generate the output stego.

Let the secret data be $\mathbf{m} = [m_1, m_2, \ldots, m_\gamma]^{\mathrm{T}} \in \{0,1\}^{\gamma}$, the quantized DCT coefficients in an JPEG image as $\mathbf{c} = [c_1, c_2, \ldots, c_{MN}]^{\mathrm{T}}$, and the stego quantized DCT coefficients as $\mathbf{y} = [y_1, y_2, \ldots, y_{MN}]^{\mathrm{T}}$. In STC, $\mathbf{m}$ is embedded into $\mathbf{c}$ using (22),

$$\mathrm{Emb}(\mathbf{c},\mathbf{m}) = \arg\min_{\mathbf{y}_l \in C(\mathbf{m})} D(\mathbf{c},\mathbf{y}) \quad (22)$$

where

$$D(\mathbf{c},\mathbf{y}) = \sum_{c_i \neq y_i} \rho_i^{y_i - c_i} \quad (23)$$

is the distortion between cover and stego image, $\mathbf{y}_l \in \{0,1\}^{MN}$ is the LSB of the quantized DCT coefficients of stego image, $C(\mathbf{m}) = \{z \in \{0,1\}^{MN} | \mathbf{H}\mathbf{z} = \mathbf{m}\}$ is the coset corresponding to $\mathbf{m}$, and $\mathbf{H} \in \{0,1\}^{\gamma \times MN}$ is a low density parity-check matrix that is pre-defined according to the embedding speed, the embedding efficiency and the payload. In **ALG.** 2, we provide the steps of data embedding.

Algorithm 2 Pseudo-Code for Data Embedding

Input: $\mathbf{m}$, $c = [c_1, c_2, \ldots, c_{MN}]^{\mathrm{T}}$
Output: Stego DCT Coefficients $c_{i,j}^{\pm}$

a. Specify the initial embedding costs $\rho_{i,j}$ of $c_{i,j}$.
b. Construct $\hat{\mathbf{I}}$ and calculate $\hat{c}_{i,j}$ using Equation (8).
c. Calculate $\beta_{i,j} = \mathrm{sign}(\hat{c}_{i,j} - c_{i,j})$ for $\hat{c}_{i,j}$, and identify α using ALG. 1.
d. Modify $\rho_{i,j}$ to $\rho_{i,j}^{+}$ and $\rho_{i,j}^{-}$ by (17) and (18).
e. Embed $\mathbf{m}$ into $\mathbf{c}$ using STC by $\rho_{i,j}^{+}$ and $\rho_{i,j}^{-}$ to obtain $c_{i,j}^{\pm}$.

In the phase of data extraction, the secret data $\mathbf{m}$ in the LSB of stego quantized DCT coefficients can be directly extracted by a matrix computation,

$$\mathbf{m} = \mathbf{H}\mathbf{y}_l \quad (24)$$

IV. EXPERIMENTAL RESULTS

To verify the proposed method, we have conducted many experiments to hide secret data. We first setup the experimental environments using the popular database. Subsequently, we analyze the quality of stego image and the variation of feature distance between a cover and a stego. Finally, we provide the results of improvements with respect to state-of-the-art works.

A. EXPERIMENT SETUP

The image dataset used in our experiments is BOSSbase ver. 1.01 [48], which contains 10000 uncompressed grayscale images sized 512×512. We compress all images using JPEG with the quality factors QF = 75 and QF = 95, respectively.

To verify the effectiveness of the proposed method, we use the popular JPEG steganographic methods of JUNIWARD, UED and UERD as benchmark. All embedding tasks are done by the STC framework. We set the payloads as 0.05, 0.1, 0.2, 0.3, 0.4, and 0.5 *bpnzac*, respectively.

Meanwhile, the most popular feature sets against JPEG steganography, i.e., ccJRM [36], DCTR [37] and GFR [38] which perform the state-of-the-art detection rate are used to check the undetectability. Since it is difficult to operate high dimensional feature sets in the SVM (support vector machine) based classification [51], [52], we employ the ensemble classifier [40] to measure the property of feature sets. The ensemble classifier consists of many FLD (Fisher Linear Discriminant) learners with low training complexity. Each FLD learner is built on a randomly selected subspace of the feature space. The final decision of the ensemble classifier is formed by aggregating the decisions of all the FLD learns. This ensemble strategy performs a comparable detectability with SVM, and achieves much low calculate complexity meanwhile. Therefore, it is reasonable to employ ensemble classifier for feature property measurement. This classification tool is also used in the works introducing ccJRM, DCTR and GFR. One half of the cover and stego feature sets are used for training, while the remaining sets are used for testing.

The criterion of evaluating the performance of feature sets is the minimal total error P_{E} with identical priors achieved on the testing sets [40],

$$P_E = \min_{P_{FA}}\left(\frac{P_{\mathrm{FA}}+P_{MD}}{2}\right) \quad (25)$$

where P_{FA} is the false alarm rate and P_{MD} the missed detection rate. The performance is evaluated using the average of P_{E} over ten random tests.

B. IMAGE QUALITY

Since the proposed method improves the distortion function, we use the functions defined in JUNIWARD, UED, and UERD, as the initial distortion functions. After improving the costs in each method, the new distortion functions for steganography were generated. We name these improved embedding using feature set ccJRM as JUNIWARD-ccJRM, UED-ccJRM, and UERD-ccJRM, respectively.

TABLE 4. Testing errors of JUNIWARD, UED, UERD and the improved version against ccJRM, DCTR, GFR and ensemble classifier.

Algorithm	Feature set	Payload (bpnzac) (QF=75)						Payload (bpnzac) (QF=95)					
		0.05	0.1	0.2	0.3	0.4	0.5	0.05	0.1	0.2	0.3	0.4	0.5
JUNIWARD	ccJRM	0.4866	0.4654	0.3951	0.3142	0.2326	0.1593	0.4993	0.4975	0.4742	0.4426	0.3925	0.3293
	DCTR	0.4737	0.4278	0.3221	0.2174	0.1336	0.0727	0.4974	0.4847	0.4452	0.3870	0.3132	0.2341
	GFR	0.4582	0.3954	0.2630	0.1530	0.0819	0.0424	0.4911	0.4717	0.4124	0.3368	0.2524	0.1739
JUNIWARD_ccJRM	ccJRM	**0.4958**	**0.4873**	**0.4496**	**0.4060**	**0.3435**	**0.2898**	**0.4975**	**0.4960**	**0.4904**	**0.4717**	**0.4456**	**0.4077**
JUNIWARD_DCTR	DCTR	**0.4909**	**0.4712**	**0.4279**	**0.3749**	**0.3073**	**0.2391**	**0.5003**	**0.4951**	**0.4957**	**0.4537**	**0.4184**	**0.3622**
JUNIWARD_GFR	GFR	**0.4789**	**0.4502**	**0.3885**	**0.3248**	**0.2622**	**0.2008**	**0.4950**	**0.4869**	**0.4568**	**0.4210**	**0.3783**	**0.3307**
UED	ccJRM	0.4576	0.4006	0.2649	0.1355	0.0501	0.0129	0.4932	0.4819	0.4340	0.3574	0.2478	0.1331
	DCTR	0.4303	0.3502	0.2052	0.1061	0.0473	0.0151	0.4824	0.4566	0.3874	0.3057	0.2144	0.1270
	GFR	0.4026	0.3008	0.1491	0.0663	0.0280	0.0105	0.4686	0.4270	0.3322	0.2358	0.1519	0.0832
UED_ccJRM	ccJRM	**0.4682**	**0.4333**	**0.3336**	**0.2177**	**0.1112**	**0.0433**	**0.4942**	**0.4859**	**0.4545**	**0.3866**	**0.2808**	**0.1591**
UED_DCTR	DCTR	**0.4503**	**0.4020**	**0.3062**	**0.2085**	**0.1108**	**0.0443**	**0.4865**	**0.4672**	**0.4232**	**0.3661**	**0.2821**	**0.1831**
UED_GFR	GFR	**0.4333**	**0.3681**	**0.2546**	**0.1679**	**0.0960**	**0.0504**	**0.4771**	**0.4483**	**0.3873**	**0.3173**	**0.2509**	**0.1769**
UERD	ccJRM	0.4825	0.4508	0.3710	0.2830	0.2060	0.1382	0.4975	0.4870	0.4567	0.4061	0.3406	0.2643
	DCTR	0.4666	0.4185	0.3089	0.2052	0.1219	0.0684	0.4932	0.4751	0.4238	0.3550	0.2787	0.1998
	GFR	0.4503	0.3812	0.2526	0.1519	0.0831	0.0453	0.4808	0.4553	0.3836	0.3010	0.2237	0.1541
UERD_ccJRM	ccJRM	**0.4892**	**0.4715**	**0.4133**	**0.3518**	**0.2821**	**0.2192**	**0.4969**	**0.4927**	**0.4715**	**0.4327**	**0.3874**	**0.3289**
UERD_DCTR	DCTR	**0.4803**	**0.4539**	**0.3954**	**0.3315**	**0.2592**	**0.1870**	**0.4944**	**0.4839**	**0.4547**	**0.4096**	**0.3541**	**0.2851**
UERD_GFR	GFR	**0.4704**	**0.4330**	**0.3576**	**0.2837**	**0.2138**	**0.1500**	**0.4892**	**0.4678**	**0.4260**	**0.3776**	**0.3195**	**0.2523**

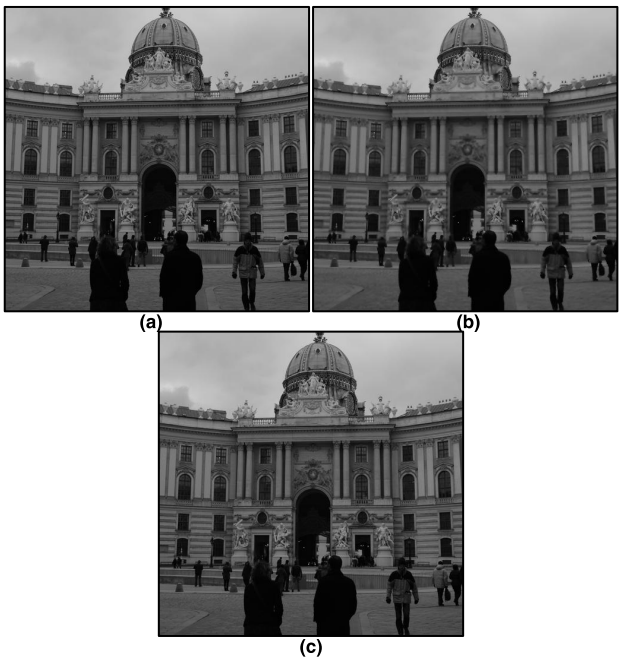

FIGURE 4. Images produced during steganography with the proposed method: (a) the cover image, (b) the reference image, (c) the stego.

Similarly, JUNIWARD-DCTR, UED-DCTR, UERD-DCTR for DCTR, and JUNIWARD-GFR, UED-GFR, UERD-GFR for GFR respectively.

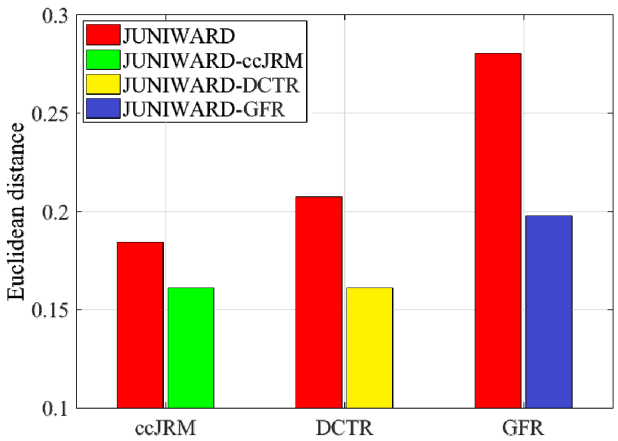

FIGURE 5. Comparison of Feature distance.

Fig. 4(a) shows a JPEG image compressed by QF = 75, and Fig. 4(b) shows the generated reference image using the proposed method. We further use JUNIWARD-DCTR to embed a secret payload of 0.5 *bpnzac* into the cover. The stego image is shown in Fig. 4(c). The PSNR value of the stego image is 46.5 dB, which indicates that the stego image is close to the cover image and so that the stego image preserves good quality even high payloads are carried.

C. COMPARISONS OF FEATURE DISTANCE

The feature distance between the cover and the stego is important for steganography security. A group of experiments

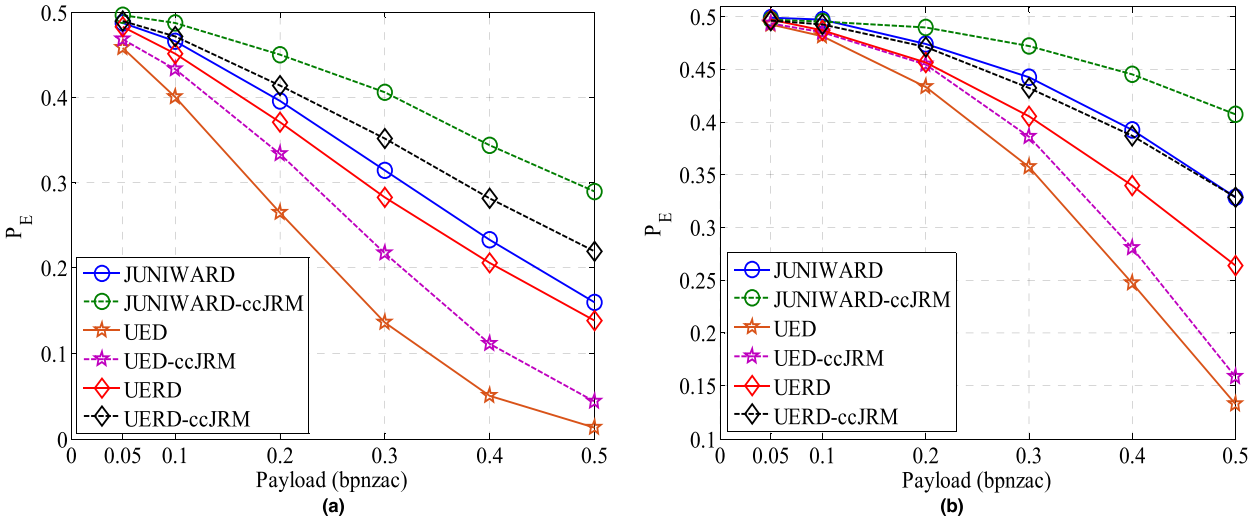

FIGURE 6. Comparisons of JUNIWARD, UED, UERD and the improved versions against ccJRM and ensemble classifier for (a) QF = 75, (b) QF = 95.

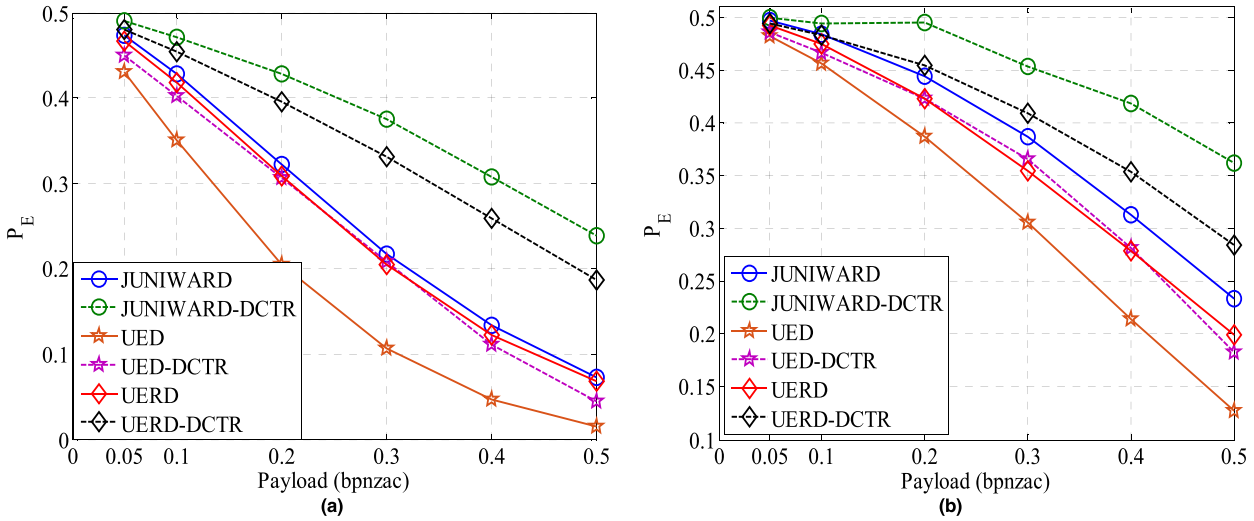

FIGURE 7. Comparisons of JUNIWARD, UED, UERD and the improved versions against DCTR and ensemble classifier for (a) QF = 75, (b) QF = 95.

are carried out on 1000 grayscale images arbitrarily selected from BOSSbase ver. 1.01. These images are compressed into JPEG with QF = 75. We use JUNIWARD and the improved JUNIWARD-ccJRM, JUNIWARD-DCTR, JUNIWARD-GFR to embed the payloads of 0.5 *bpnzac* into the JPEG images, respectively.

Fig. 5 shows the average Euclidean distances between the cover and stego images, using the feature sets defined by ccJRM, DCTR, and GFR. The results indicate that the feature distances are efficiently decreased after improving the costs of the distortion function. Once the feature distances are shortened, the undetectability of steganography would increase. Next, we discussed the capability of anti-steganalysis.

D. UNDETECTABILITY AGAINST STEGANALYSIS

We employ all the 10000 images in BOSSbase ver. 1.01, and compress them by JPEG using the quality factors 75 and 95, respectively. We first implement the data embedding by JUNIWARD, UED, and UERD. Next, we use the proposed approach to improve these methods to JUNIWARD-ccJRM, UED-ccJRM, UERD-ccJRM, JUNIWARD-DCTR, UED- DCTR, UERD-DCTR, JUNIWARD-GFR, UED-GFR, UERD- GFR, which are then used to embed data into the covers. To evaluate the security of the proposed method, we use ccJRM, DCTR, and GFR to analyze the undetectability of the stego images.

Fig. 6 $\sim$ Fig. 8 show the undetectability comparisons of these methods when the feature set used for steganalysis is

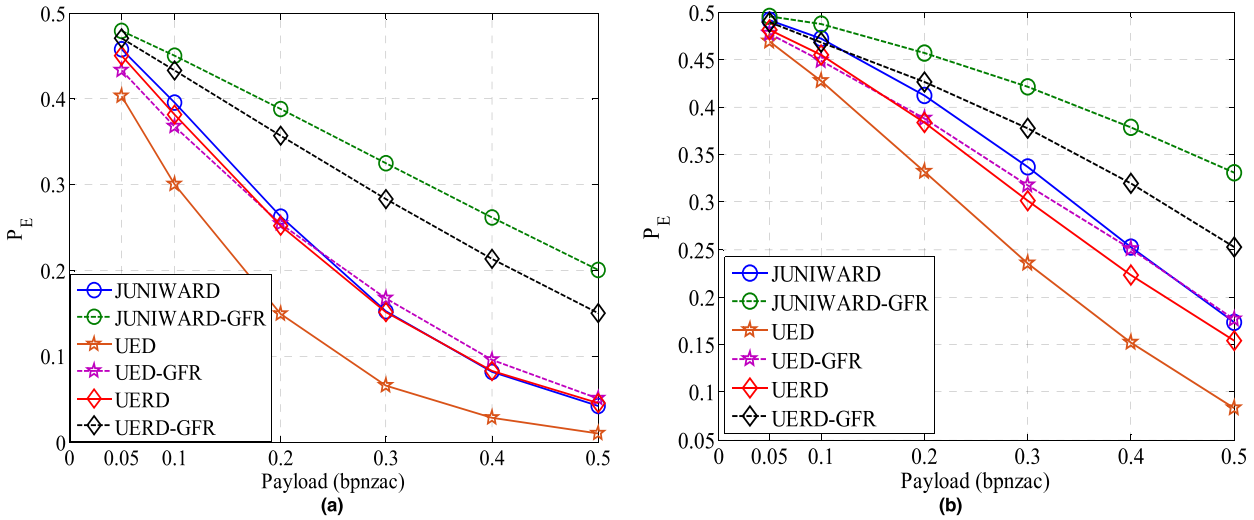

FIGURE 8. Comparisons of JUNIWARD, UED, UERD and the improved versions against GFR and ensemble classifier for (a) QF = 75, (b) QF = 95.

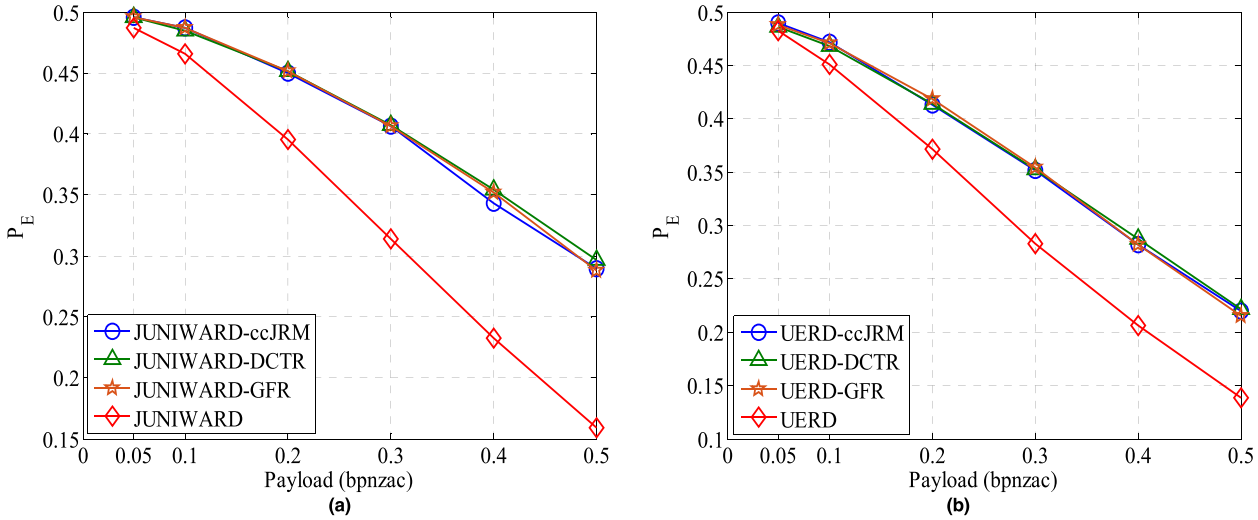

FIGURE 9. Mismatching comparisons against ccJRM and ensemble classifier for QF = 75 with (a) JUNIWARD, (b) UERD.

matched with the one for embedding, and Table 4 depicts all numerical values. The results indicate that the security performances of all approaches for JPEG steganography are improved by using the proposed method.

With the improving method, the P_E for JUNIWARD is improved by 16.64% for 0.5 bpnzac and QF = 75 against DCTR, 18.03% for 0.4 bpnzac and QF = 75 against GFR, and 15.68% for 0.5 bpnzac and QF = 95 against GFR.

For UED, the improvement is 10.24% for 0.3 bpnzac and QF = 75 against DCTR, 10.55% for 0.2 bpnzac and QF = 75 against GFR, and 9.90% for 0.4 bpnzac and QF = 95 against GFR.

For UERD, the improvement is 13.73% for 0.4 bpnzac and QF = 75 against DCTR, 13.18% for 0.3 bpnzac and QF = 75, 9.82% for 0.5 bpnzac and QF = 95 against GFR.

The average improvement on P_E of JUNIWARD tested on image with QF = 75 is 9.96 %, 7.54% for UERD, and 5.92% for UED. For QF = 95, the average improvement is 5.37% for JUNIWARD, 4.15% for UERD, and 3.93% for UED.

The results indicate that the covers with a smaller quality factor can result in a better improvement on P_E. For a higher quality factor, the detection error P_E of steganalytic tools is closer to 50%, i.e., the bound of P_E. Therefore, the improving space is limited for the JPEG covers with higher quality factors. Meanwhile, when few payloads are embedded into the cover, it is difficult to further increase detection error P_E, since P_E is already around 50%.

However, it is difficult to achieve the matching of feature set in practice. In consideration of this, as shown in Fig. 9 ∼ Fig. 11, we conducted some undetectability

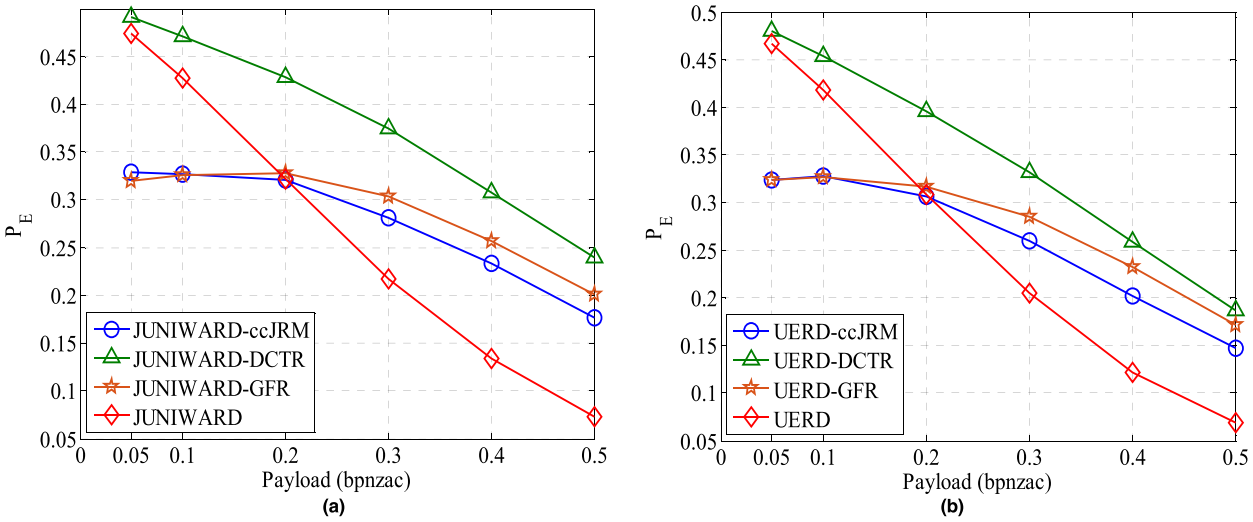

FIGURE 10. Mismatching comparisons against DCTR and ensemble classifier for QF = 75 with (a) JUNIWARD, (b) UERD.

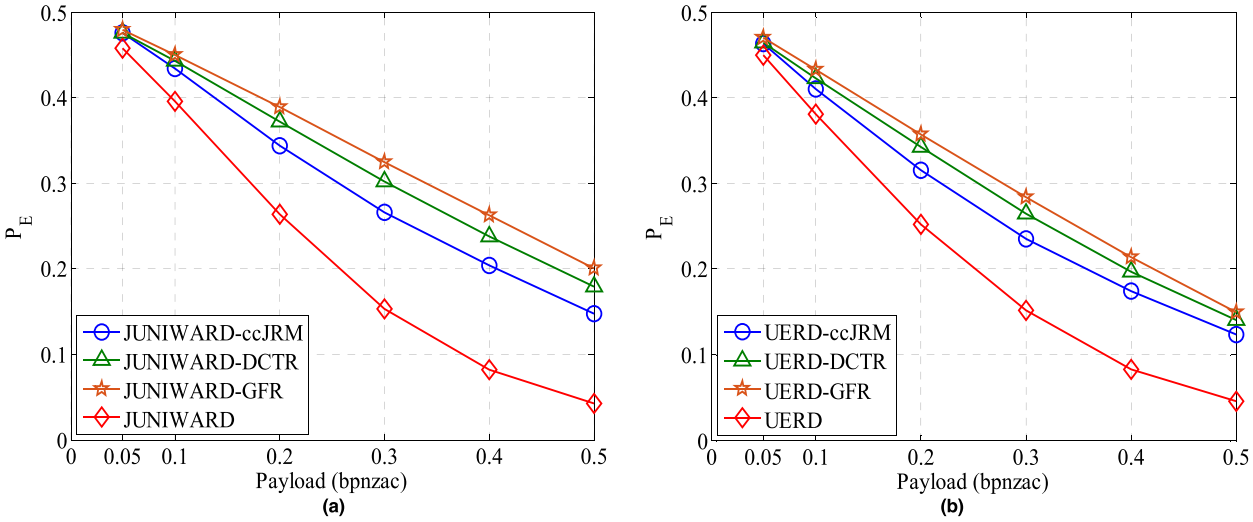

FIGURE 11. Mismatching comparisons against GFR and ensemble classifier for QF = 75 with (a) JUNIWARD, (b) UERD.

comparisons for the cases of feature set mismatching. To observe the improvement on P_E easier, we conducted the mismatching comparisons for QF = 75.

It can be seen from Fig. 9 that the three feature sets ccJRM, DCTR and GFR guided improving embedding perform almost same undetectability against ccJRM. However, ccJRM or GFR guided improving embedding perform obvious defects in low payloads against DCTR, as shown in Fig. 10. To resist GFR, the GFR guided improving embedding performs the best. But DCTR guided improving embedding also achieves a comparable undetectability, as shown in Fig. 11.

In total, the DCTR guided improving embedding performs satisfactory undetectability. Furthermore, the computation complexity of DCTR is lower comparing with ccJRM [36] and GFR [37]. Therefore, we recommend DCTR as a universal feature set to guide the improving of distortion functions. That means the universal α in Equations (15) and (16) is obtained by the guidance of DCTR.

Finally, we compare the proposed method with [45], which is also a kind of distortion improving based steganography. The experiments are conducted using the image set BURST proposed in [45]. The database contains 9310 cover images and 9310 reference images, each of which is sized 512×512. We compress 9310 cover images by JPEG using the quality factors QF = 75. Fig. 12 shows the undetectability comparison results using the steganalytic tools of ccJRM and DCTR, respectively. The "JUNIWARD-[45]" stands for the improved JUNIWARD using the method in [45]. In the pioneering method in [45], an outstanding increment on security performance is achieved using multiple images sharing the same scene as references. As shown in Fig. 12, it is clear

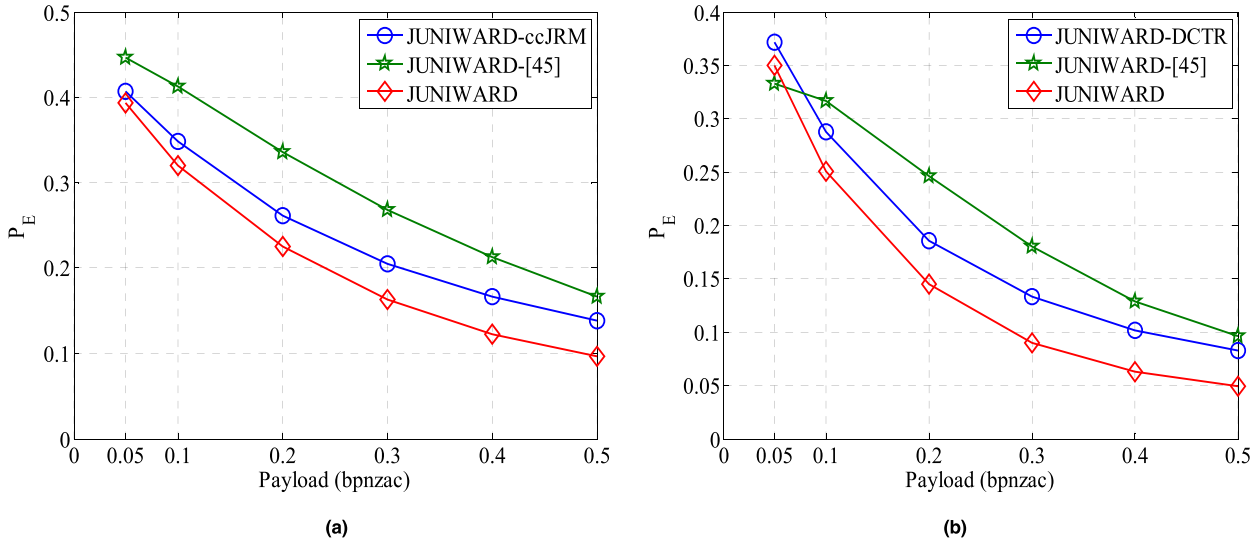

FIGURE 12. Comparing the proposed method with [45] for JUNIWARD on BURST using QF = 75 against (a) ccJRM, (b) DCTR.

that the performance of the method in [45] indicates better anti-steganalysis efficiency than the proposed method. Nevertheless, the method in [45] requires a data hider to provide multiple images sharing the same scene to implement the improvement. In many applications, it is difficult for a data hider to supply the similar images. In the proposed method, only one image is required and the improving is realized by constructing a reference image.

V. CONCLUSION

In this paper we propose a framework to improve security performance of existing JPEG steganographic method by shortening the distortion in feature space of steganalysis. We first construct a reference image by Wiener filtering. Guided by the reference image, we determine the direction of $+1$ or -1 when modifying the DCT coefficients. With the public knowledge of steganalysis, the modification extensions are identified by evaluating the feature variations before and after data embedding.

Experimental results show that the security performance of proposed method outperforms state-of-the-art steganography method for JPEG images. The proposed algorithm is universal for improving all distortion functions for JPEG steganography proposed in previous arts, e.g., JUNIWARD, UED, and UERD. For further study we will focus on the implementations of steganography guided by more public knowledge e.g. multiple images, and general improving of the distortion functions for both the spatial and JPEG images.

REFERENCES

[1] Z. Wang, X. Zhang, and Z. Yin, "Joint cover-selection and payload-allocation by steganographic distortion optimization," *IEEE Signal Process. Lett.*, vol. 25, no. 10, pp. 1530–1534, Oct. 2018.

[2] F. Li, K. Wu, X. Zhang, J. Yu, J. Lei, and M. Wen, "Robust batch steganography in social networks with non-uniform payload and data decomposition," *IEEE Access*, vol. 6, pp. 29912–29925, 2018.

[3] D. Upham. *JSTEG Code*. Accessed: 1993. [Online]. Available: http://zooid.org/~paul/crypto/jsteg

[4] J. Fridrich and M. Goljan, "Practical steganalysis of digital images: State of the art," *Proc. SPIE*, vol. 4675, pp. 1–13, Apr. 2002.

[5] H. Wang and S. Wang, "Cyber warfare: Steganography vs. steganalysis," *Commun. ACM*, vol. 47, no. 10, pp. 76–82, Oct. 2004.

[6] A. D. Ker, "Improved detection of LSB steganography in grayscale images," in *Proc. Int. Workshop Inf. Hiding*, Toronto, ON, Canada, May 2004, pp. 97–115.

[7] A. Westfeld, "F5—A steganographic algorithm," in *Proc. 4th Int. Workshop Inf. Hiding*, Pittsburgh, PA, USA, Apr. 2001, pp. 289–302.

[8] S. Phil, "Model-based steganography," in *Proc. 2nd Int. Workshop Digit.-Forensics Watermarking*, Seoul, South Korea, Oct. 2003, pp. 154–167.

[9] J. Fridrich, T. Pevný, and J. Kodovský, "Statistically undetectable jpeg steganography: Dead ends challenges, and opportunities," in *Proc. 9th Workshop Multimedia Secur.*, New York, NY, USA, Sep. 2007, pp. 3–14.

[10] A. D. Ker, "Steganalysis of LSB matching in grayscale images," *IEEE Signal Process. Lett.*, vol. 12, no. 6, pp. 441–444, Jun. 2005.

[11] F. Huang, B. Li, and J. Huang, "Attack LSB matching steganography by counting alteration rate of the number of neighbourhood gray levels," in *Proc. 14th IEEE Int. Conf. Image Process. (ICIP)*, Paris, France, Oct. 2007, pp. 401–404.

[12] J. Zhang and D. Zhang, "Detection of LSB matching steganography in decompressed images," *IEEE Signal Process. Lett.*, vol. 17, no. 2, pp. 141–144, Feb. 2010.

[13] J. Fridrich and D. Soukal, "Matrix embedding for large payloads," *IEEE Trans. Inf. Forensics Security*, vol. 1, no. 3, pp. 390–395, Sep. 2006.

[14] X. Zhang and S. Wang, "Efficient steganographic embedding by exploiting modification direction," *IEEE Commun. Lett.*, vol. 10, no. 11, pp. 781–783, Nov. 2006.

[15] W. Zhang, X. Zhang, and S. Wang, "Maximizing steganographic embedding efficiency by combining hamming codes and wet paper codes," in *Proc. 10th Int. Workshop Inf. Hiding*, Santa Barbara, CA, USA, May 2008, pp. 60–71.

[16] T. Filler, J. Judas, and J. Fridrich, "Minimizing additive distortion in steganography using syndrome-trellis codes," *IEEE Trans. Inf. Forensics Security*, vol. 6, no. 3, pp. 920–935, Sep. 2011.

[17] V. Holub and J. Fridrich, "Designing steganographic distortion using directional filters," in *Proc. IEEE Int. Workshop Inf. Forensics Secur.*, Binghamton, NY, USA, Dec. 2012, pp. 234–239.

[18] V. Holub and J. Fridrich, "Digital image steganography using universal distortion," in *Proc. 1st ACM Workshop Inf. Hiding Multimedia Secur.*, New York, NY, USA, Jun. 2013, pp. 59–68.

[19] V. Holub, J. Fridrich, and T. Denemark, "Universal distortion function for steganography in an arbitrary domain," *EURASIP J. Inf. Secur.*, vol. 2014, no. 1, pp. 1–13, 2014.

[20] B. Li, M. Wang, J. Huang, and X. Li, "A new cost function for spatial image steganography," in *Proc. IEEE Int. Conf. Image Process.*, Paris, France, Oct. 2014, pp. 4206–4210.
[21] V. Sedighi, R. Cogranne, and J. Fridrich, "Content-adaptive steganography by minimizing statistical detectability," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 2, pp. 221–234, Feb. 2016.
[22] S. Kouider, M. Chaumont, and W. Puech, "Adaptive steganography by oracle (ASO)," in *Proc. IEEE Int. Conf. Multimedia Expo*, San Jose, CA, USA, Jul. 2013, pp. 1–6.
[23] L. Guo, J. Ni, and Y. Q. Shi, "Uniform embedding for efficient JPEG steganography," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 5, pp. 814–825, May 2014.
[24] L. Guo, J. Ni, W. Su, C. Tang, and Y.-Q. Shi, "Using statistical image model for JPEG steganography: Uniform embedding revisited," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 12, pp. 2669–2680, Dec. 2015.
[25] Y. Pan, J. Ni, and W. Su, "Improved uniform embedding for efficient JPEG steganography," in *Proc. Int. Conf. Cloud Comput. Secur.*, Nanjing, China, Jun. 2016, pp. 125–133.
[26] Z. Wang, X. Zhang, and Z. Yin, "Hybrid distortion function for JPEG steganography," *J. Electron. Imag.*, vol. 25, no. 5, p. 050501, Sep. 2016.
[27] Q. Wei, Z. Yin, Z. Wang, and X. Zhang, "Distortion function based on residual blocks for JPEG steganography," *Multimedia Tools Appl.*, vol. 77, no. 14, pp. 17875–17888, Jul. 2018.
[28] J. Fridrich and J. Kodovský, "Rich models for steganalysis of digital images," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 3, pp. 868–882, Jun. 2012.
[29] V. Holub and J. Fridrich, "Random projections of residuals for digital image steganalysis," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 12, pp. 1996–2006, Dec. 2013.
[30] T. Denemark, V. Sedighi, V. Holub, R. Cogranne, and J. Fridrich, "Selection-channel-aware rich model for Steganalysis of digital images," in *Proc. IEEE Int. Workshop Inf. Forensics Secur.*, Atlanta, GA, USA, Dec. 2014, pp. 48–53.
[31] V. Holub and J. Fridrich, "Phase-aware projection model for steganalysis of JPEG images," *Proc. SPIE*, vol. 9409, pp. 94090T-1–94090T-11, Mar. 2015.
[32] M. Chen, V. Sedighi, M. M. Boroumand, and J. Fridrich, "JPEG-phase-aware convolutional neural network for steganalysis of JPEG images," in *Proc. 5th ACM Workshop Inf. Hiding Multimedia Secur.*, Philadelphia, PA, USA, Jun. 2017, pp. 75–84.
[33] G. Xu, "Deep convolutional neural network to detect J-UNIWARD," in *Proc. 5th ACM Workshop Inf. Hiding Multimedia Secur.*, Philadelphia, PA, USA, Jun. 2017, pp. 67–73.
[34] J. Zeng, S. Tan, B. Li, and J. Huang, "Large-scale JPEG image steganalysis using hybrid deep-learning framework," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 5, pp. 1200–1214, May 2018.
[35] B. Li, Z. Li, S. Zhou, S. Tan, and X. Zhang, "New steganalytic features for spatial image steganography based on derivative filters and threshold LBP operator," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 5, pp. 1242–1257, May 2018.
[36] J. Kodovský and J. Fridrich, "Steganalysis of JPEG images using rich models," *Proc. SPIE*, vol. 8303, pp. 83030A-1–83030A-13, Jan. 2012.
[37] V. Holub and J. Fridrich, "Low-complexity features for JPEG steganalysis using undecimated DCT," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 2, pp. 219–228, Feb. 2014.
[38] X. Song, F. Liu, C. Yang, X. Luo, and Y. Zhang, "Steganalysis of adaptive JPEG steganography using 2D Gabor filters," in *Proc. 3rd ACM Workshop Inf. Hiding Multimedia Secur.*, New York, NY, USA, Jun. 2015, pp. 15–23.
[39] C. Xia, Q. Guan, X. Zhao, Z. Xu, and Y. Ma, "Improving GFR steganalysis features by using Gabor symmetry and weighted histograms," in *Proc. 5th ACM Workshop Inf. Hiding Multimedia Secur.*, Philadelphia, PA, USA, Jun. 2017, pp. 55–66.
[40] J. Kodovský, J. Fridrich, and V. Holub, "Ensemble classifiers for steganalysis of digital media," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 2, pp. 432–444, Apr. 2012.
[41] F. Li, X. Zhang, B. Chen, and G. Feng, "JPEG steganalysis with high-dimensional features and Bayesian ensemble classifier," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 233–236, Mar. 2013.
[42] J. Kodovský and J. Fridrich, "On completeness of feature spaces in blind steganalysis," in *Proc. 10th ACM Workshop Multimedia Secur.*, Oxford, U.K., Sep. 2008, pp. 123–132.
[43] T. Denemark and J. Fridrich, "Side-informed steganography with additive distortion," in *Proc. IEEE Int. Workshop Inf. Forensics Secur.*, Rome, Italy, Nov. 2015, pp. 16–19.
[44] T. Denemark and J. Fridrich, "Model based steganography with precover," *Electron. Imag.*, vol. 11, pp. 56–66, Jan. 2017.
[45] T. Denemark and J. Fridrich, "Steganography with multiple JPEG images of the same scene," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 10, pp. 2308–2319, Oct. 2017.
[46] B. Li, T. T. Ng, X. Li, S. Tan, and J. Huang, "Statistical model of JPEG noises and its application in quantization step estimation," *IEEE Trans. Image Process.*, vol. 24, no. 5, pp. 1471–1484, May 2015.
[47] M. Kutter and S. Winkler, "A vision-based masking model for spread-spectrum image watermarking," *IEEE Trans. Image Process.*, vol. 11, no. 1, pp. 16–25, Jan. 2002.
[48] P. Bas, T. Filler, and T. Pevný, "'Break our steganographic system': The ins and outs of organizing BOSS," in *Proc. 13th Int. Conf. Inf. Hiding*, Prague, Czech Republic, May 2011, pp. 59–70.
[49] Z. Wang, J. Lv, Q. Wei, and X. Zhang, "Distortion Function for Spatial Image Steganography Based on the Polarity of Embedding Change," in *Proc. 15th Int. Workshop Digit.-Forensics Watermarking (IWDW)*, Beijing, China, Sep. 2016, pp. 487–493.
[50] J. Fridrich and T. Filler, "Practical methods for minimizing embedding impact in steganography," *Proc. SPIE*, vol. 6505, pp. 650502-2–650502-3, Feb. 2007.
[51] T. Pevný and J. Fridrich, "Novelty detection in blind steganalysis," in *Proc. 10th ACM Workshop Multimedia Secur.*, New York, NY, USA, Sep. 2008, pp. 167–176.
[52] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 27:1–27:27, 2011.

ZICHI WANG received the B.S. degree in electronics and information engineering and the M.S. degree in signal and information processing from Shanghai University, China, in 2014 and 2017, respectively, where he is currently pursuing the Ph.D. degree. His research interests include steganography, steganalysis, and reversible data hiding. He has published about 20 papers in these areas.

ZHENXING QIAN received the B.S. and Ph.D. degrees from the University of Science and Technology of China in 2003 and 2007, respectively. Since 2009, he has been a Faculty Member with the School of Communication and Information Engineering, Shanghai University, where he is currently a Professor. His research interests include information hiding and multimedia security.

XINPENG ZHANG received the B.S. degree in computational mathematics from Jilin University, China, in 1995, and the M.E. and Ph.D. degrees in communication and information system from Shanghai University, China, in 2001 and 2004, respectively. Since 2004, he has been a Faculty Member with the School of Communication and Information Engineering, Shanghai University, where he is currently a Professor. His research interests include information hiding, image processing, and digital forensics. He has published over 200 papers in these areas.

MIN YANG received the B.Sc. and Ph.D. degrees in computer science from Fudan University in 2001 and 2006, respectively. He is currently a Professor with the Software School, Fudan University. His research interests are in system software and security.

DENGPAN YE was born in Hubei, China. He received the B.A.Sc. degree in automatic control from SCUT in 1996 and the Ph.D. degree from NJUST in 2005. He was a Post-Doctoral Fellow with the Information System School, Singapore Management University. Since 2012, he has been a Professor with the School of Computer Science, Wuhan University. He has authored or co-authored more than 30 refereed journal and conference papers. His research interests include machine learning and multimedia security.