Secure Halftone Image Steganography Based on Feature Space and Layer Embedding

Wei Lu, *Member, IEEE*, Junjia Chen, Junhong Zhang, Jiwu Huang, *Fellow, IEEE*, Jian Weng, *Member, IEEE*, and Yicong Zhou, *Senior Member, IEEE*

***Abstract*—Syndrome-trellis codes (STCs) are commonly used in image steganographic schemes, which aim at minimizing the embedding distortion, but most distortion models cannot capture the mutual interaction of embedding modifications (MIEMs). In this article, a secure halftone image steganographic scheme based on a feature space and layer embedding is proposed. First, a feature space is constructed by a characterization method that is designed based on the statistics of 4 × 4 pixel blocks in halftone images. Upon the feature space, a generalized steganalyzer with good classification ability is proposed, which is used to measure the embedding distortion. As a result, a distortion model based on a hybrid feature space is constructed, which outperforms some state-of-the-art models. Then, as the distortion model is established on the statistics of local regions, a layer embedding strategy is proposed to reduce MIEM. It divides the host image into multiple layers according to their relative positions in 4 × 4 blocks, and the embedding procedure is executed layer by layer. In each layer, any two pixels are located at different 4 × 4 blocks in the original image, and the distortion model makes sure that the calculation of pixel distortions is independent. Between layers, the pixel distortions of the current layer are updated according to the previous embedding modifications, thus reducing the total embedding distortion. Comparisons with prior schemes demonstrate that the proposed steganographic scheme achieves high statistical security when resisting the state-of-the-art steganalysis.**

***Index Terms*—Data hiding, feature space, halftone image, steganography, syndrome-trellis codes (STCs).**

Manuscript received March 15, 2020; revised July 15, 2020; accepted September 20, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant U1736118, Grant 62072480, and Grant U19B2022; in part by the Key Areas Research and Development Program of Guangdong under Grant 2019B010136002 and Grant 2019B010139003; in part by the Key Project of Scientific Research Plan of Guangzhou under Grant 201804020068; and in part by the Shenzhen Research and Development Program under Grant GJHZ20180928155814437. This article was recommended by Associate Editor Y. Yuan. *(Corresponding author: Wei Lu.)*

Wei Lu, Junjia Chen, and Junhong Zhang are with the School of Data and Computer Science, Guangdong Province Key Laboratory of Information Security Technology, Ministry of Education Key Laboratory of Machine Intelligence and Advanced Computing, Sun Yat-sen University, Guangzhou 510006, China (e-mail: luwei3@mail.sysu.edu.cn; chenjj233@mail2.sysu.edu.cn; zhangjh65@mail2.sysu.edu.cn).

Jiwu Huang is with the Guangdong Key Laboratory of Intelligent Information Processing, Shenzhen Key Laboratory of Media Security, Guangdong Laboratory of Artificial Intelligence and Digital Economy, and the Shenzhen Institute of Artificial Intelligence and Robotics for Society, Shenzhen University, Shenzhen 518060, China (e-mail: jwhuang@szu.edu.cn).

Jian Weng is with the College of Information Science and Technology and the College of Cyber Security, Jinan University, Guangzhou 510632, China (e-mail: cryptjweng@gmail.com).

Yicong Zhou is with the Department of Computer and Information Science, University of Macau, Macau, China (e-mail: yicongzhou@um.edu.mo).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCYB.2020.3026047

I. Introduction

STEGANOGRAPHY, as a branch of data hiding [1]–[3], aims at hiding secret messages into digital media and being imperceptible to others [4]–[6]. Different from watermarking techniques, which is another kind of data hiding and usually can resist linear or other transformation of digital images, steganography pays more attention to the statistical security and should provide high integrity of secret messages. As a kind of digital media, binary images only have two kinds of states: "0" and "1," which correspondingly represent the pixel color of black and white. Based on the characteristic, secret messages are hidden by pixel toggling. Binary images can be roughly divided into two kinds, including: 1) halftone images and 2) ordinary binary images. Different from ordinary binary images, halftone images (two-tone images) are usually transferred from grayscale images (multitone images) by halftoning techniques, including error diffusion [7], dithering [8], dot diffusion [9], direct binary search [10], etc.

In previous works, many existing halftone image data hiding schemes focused on designing a distortion model to select "slave" pixels after hiding one bit at each pseudorandom position at which the pixels located are regarded as "master" pixels [11], [12]. Data hiding by smart pair toggling (DHSPT) was proposed by Fu and Au [11] in which a distortion model is defined to measure the connectivity of master pixels with neighboring slave pixels. The slave pixel is selected with the minimum connectivity value. Pair toggling with the human visual system (PTHVS) was proposed by Guo [12] to improve the selection of slave pixels. Guo proposed a distortion model based on the human visual system, which improves the weights of pixel connectivity along with the horizontal, vertical, and diagonal directions. However, the master pixels at the pseudorandom positions still cause "salt-and-pepper" clusters, and the visual quality is destroyed when the secret message's length increases and the security performance decreases.

In ordinary binary image steganography, many schemes proposed various embedding distortion models to evaluate the pixel distortions, which are different from the strategy of "master–slave" pixel toggling. Some research on those is based on statistics [13], [14]. Feng *et al.* [13] proposed a flipping

distortion model (FDM) and the alternation of the patterns' distributions are employed to indicate the pixel flippability. Yeung *et al.* [14] presented a statistical prediction distortion model (PDM) and introduced the concept of "uncertainty" to evaluate the pixel distortions. Although the image expression between ordinary binary images and halftone images is different, the statistics-based FDM and PDM can be used to evaluate the embedding distortion for halftone images.

In addition to the establishment of embedding distortion models, which is an important step in steganographic schemes, different embedding strategies have a huge impact on the statistical security of steganography [13]–[17]. Feng *et al.* [13] and Lu *et al.* [15] used superpixels as the cover vector to embed the message segment by applying the syndrome-trellis codes (STC) [18] encoder. Yeung *et al.* [14] demonstrated that using single pixels to apply the STC encoder outperforms using superpixels as STC's carriers in terms of the average embedding distortion. However, these strategies do not consider the mutual interaction of embedding modifications (MIEMs). Thus, they easily toggle adjacent pixels simultaneously and enhance the risk of detection by the steganalyzer.

To exploit MIEM and improve the statistical security, synchronizing the selection channel [17] and clustering modification directions [16] were independently proposed for grayscale image steganography. In their strategies, adjacent pixels are divided into two nonoverlapped subsets. The pixel distortions of the second subset are updated according to the changes of neighboring pixels. Furthermore, Zhang *et al.* [5] defined a joint distortion on pixel blocks to exploit MIEM and decomposed it for minimizing nonadditive distortion with low computational complexity. However, the schemes [5], [16], [17] do not consider the statistical region of the initial distortion model in the embedding procedure. For instance, Zhang *et al.* used HILL [19] as their initial distortion whose cost function should select three filters. One of those filters has the default size of 15×15, which is larger than the block they used to define their joint distortion. Therefore, the mutual interaction among the changed pixels cannot be avoided.

This article proposes an embedding distortion model defined in the feature space to accurately evaluate the embedding distortion. First, a feature space is constructed by a characterization method which is the statistics of 4×4 pixel blocks. Then, a generalized steganalyzer with good classification ability is proposed for designing a better distortion model. Finally, we construct a combined distortion model based on a hybrid feature space to resist various steganalysis. Consequently, experimental results demonstrate that the proposed distortion model effectively evaluates the embedding distortion and outperforms the models presented in [13] and [14].

Furthermore, a layer embedding strategy based on 4×4 blocks is proposed to reduce MIEM and improve statistic security. It divides the host image into multiple layers and the embedding procedure executes layer by layer. Any two pixels in the same layer are located at different 4×4 blocks but their positions in the block are the same and not affected by other pixels. That is, the embedding modifications can avoid mutually interacting in the same layer. Besides, the pixel value of the current layer is updated according to the changes of neighboring pixels in the previous layer embedding. As a result, the subsequent embedding positions are optimized by the STC encoder [18]. Experimental results show that the proposed layer embedding strategy can reduce the total distortion and improve statistical security.

In summary, the main contributions of the proposed scheme are as follows.

1) A distortion measurement based on the feature space is proposed. Upon it, a generalized steganalyzer with a good classification ability is used as a guide for designing the distortion model.
2) A diversified distortion model is proposed. The model is based on a hybrid feature space and the experimental results demonstrate that it evaluates the embedding distortion accurately.
3) A layer embedding strategy is proposed to reduce MIEM and the embedding modifications do not mutually interact in each layer. Furthermore, the pixel distortions of the current layer are updated according to the previous embedding modifications, which reduces the total distortion.

The remainder of this article is organized as follows. In Section II, we detail the construction of the embedding distortion models, including a basic model and a combined model. In Section III, the layer embedding strategy is elaborated. Section IV shows the entire proposed steganographic scheme. Section V presents the experimental results and discussion. Finally, Section VI concludes this article.

II. Embedding Distortion Model

Due to the principle of minimal impact embedding [20], the design of steganographic schemes can be decomposed into the investigation of the image model and coder. In terms of the coder designs, STC [18] is a practical method to embed messages that can approach the lower bound of average distortion. Beyond that, a better design of the image model can also improve the steganographic scheme to achieve higher statistical security. In this section, we focus on designing the embedding distortion model based on a feature space.

A. Basic Model Based on Feature Space

A specific characterization method can map objects into a specific feature space. Consequently, an object is represented as a specific feature vector. In the field of image steganalysis [21]–[24], the design of the characterization method is an important step. They use an appropriate characterization method to map the cover and stego images into the feature space so that different categories of images can be clearly identified. As an adversary of steganalysis, many steganographic schemes [13], [23] propose a distortion function defined as the weighted norm of the difference between feature vectors of the cover and stego images in the chosen feature space.

Although many steganographic schemes have proposed the distortion model with the same type, applying the feature

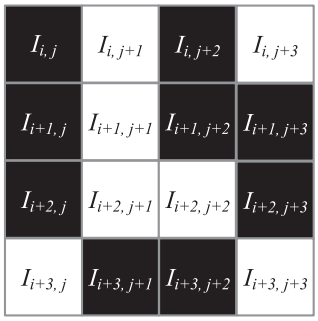


Fig. 1. Example of 4×4 block, whose $t = T(I_{i,j}) = 2^1 + 2^3 + 2^5 + 2^9 + 2^{10} + 2^{12} + 2^{15} = 38442$. $I_{i,j}$ represents the value of the first pixel in the block.

spaces to halftone images directly does not yield satisfactory results. HUGO [23] preserves the decomposed SPAM [24] model, but the SPAM feature is proposed to detect steganographic algorithms for grayscale images. FDM [13] uses 3×3 local texture patterns to describe the relationship of the texture structure in ordinary binary images. Although halftone images are a kind of binary image, the description of halftone images is totally different from that of ordinary binary images. Ordinary binary images usually exhibit their contents by edge lines, while halftone image contents are expressed by pixel mesh density. Since the aspects of them differ, it is necessary to design a characterization method for halftone images and an embedding distortion model upon it.

Halftone images are perceived as continuous-tone images when viewed at a distance due to the low-pass filtering effect of human visual perception. To simulate the human visual perception for halftone images better, large scope of statistics is necessary. Concretely, a histogram of 4×4 pixel blocks is proposed to describe the texture structure distribution for halftone images. First, let $T(I_{i,j})$ denote a function for obtaining the type of a 4×4 block shown in Fig. 1. It can be written as

$$T\big(I_{i,j}\big) = \sum_{k=0}^{3} \sum_{l=0}^{3} 2^{4k+l} \times I_{i+k,j+l} \tag{1}$$

where $I_{i,j}$ denotes the (i,j)th pixel values in the image I and $I_{i+k,j+l} \in \{0,1\}$, $k,l \in \{0,1,2,3\}$ denotes the pixel values of the block. It is worth mentioning that the values of black and white pixels are assigned with "0" and "1," respectively. According to (1), the total number of block types reaches $2^{16} = 65536$ and $T(I_{i,j}) \in \{0,1,\ldots,65535\}$. After the function definition for obtaining the block type, H_t denotes an element with type t in the normalized histogram of the 4×4 blocks, that is

$$H_t = \frac{1}{\lambda} \sum_{i=0}^{n_1-4} \sum_{j=0}^{n_2-4} \delta\big(T\big(I_{i,j}\big) = t\big) \tag{2}$$

where $n_1 \times n_2$ is the size of images, $\delta(\bullet) = 1$ only if its argument is satisfied, otherwise $\delta(\bullet) = 0$, λ is the normalization factor ensuring that $\sum_{t=0}^{65535} H_t = 1$ and $t \in \{0,1,\ldots,65535\}$. According to (2), an image can be expressed as

$$\mathbf{v} = [H_0, H_1, \ldots, H_{65535}]. \tag{3}$$

It can be observed that the dimensionality of a feature vector is 65 536. In steganalysis, a high-dimensional feature vector easily causes the curse of dimensionality. In contrast, when designing steganographic schemes, a high-dimensional feature space is acceptable. For example, HUGO [4] observes the distribution of cover and stego images in the feature space and heuristically sets the weight for evaluating pixel toggling of each feature. However, a heuristic model fails to weight each feature accurately. This article proposes to utilize feature space constructed by cover and stego images to accurately weight each feature. To avoid the curse of dimensionality, dimensionality reduction is necessary.

Principal component analysis (PCA) [25] is a statistical procedure that uses an orthogonal transformation to reduce the dimensionality of feature sets while maximizing the variance of projected features. PCA is employed to reduce the dimensionality of the proposed histogram and the final characterization method is denoted as HPCA. To endow HPCA with a good ability to distinguish the cover and stego images, PCA maximizes the variance of feature sets $\mathbf{M}$ that consists of

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}^c \\ \mathbf{M}^s \end{bmatrix} \tag{4}$$

where $\mathbf{M}^c$ and $\mathbf{M}^s$ denote the feature sets of the cover and stego images, respectively. They can be written as

$$\begin{aligned} \mathbf{M}^c &= \left[\mathbf{v}_1^c, \mathbf{v}_2^c, \ldots, \mathbf{v}_n^c, \ldots, \mathbf{v}_N^c\right]^{\mathbf{T}} \\ \mathbf{M}^s &= \left[\mathbf{v}_1^s, \mathbf{v}_2^s, \ldots, \mathbf{v}_n^s, \ldots, \mathbf{v}_N^s\right]^{\mathbf{T}} \end{aligned} \tag{5}$$

where $\mathbf{v}_n^c$ and $\mathbf{v}_n^s$, $n \in \{1,2,\ldots,N\}$ represent the nth feature vector of the cover and simulated stego images, respectively. N is the total number of training sets in the image database. The simulated stego images are created by embedding with a simulated payload, which will be discussed in Section II-B. According to the principle of PCA, the Hotelling transformation [25] can transform the feature vector from high to low dimensionality

$$\hat{\mathbf{M}} = (\mathbf{M} - \mathbf{R})\mathbf{P} \tag{6}$$

where $\mathbf{P}$ is a reduced orthogonal transformation matrix and $\mathbf{R}$ consists of the $2N$ mean vector $\mathbf{m} = (1/2N) \sum_{n=1}^{N} (\mathbf{v}_n^c + \mathbf{v}_n^s)$. Depending on $\mathbf{P}$, the proposed feature vector $\mathbf{v}$ can be transferred to a low-dimensional vector. It is worth mentioning that in the proposed scheme, the dimensionality of feature space is reduced to 600 by PCA.

A good characterization method can be used as a guide for designing an embedding distortion model [4], [23]. Table I shows the detection performance comparisons of the different characterization methods. The experimental setup is the same as that in Section V. These methods are used to detect the stego images created by FDMS [13] and PDMS [14]. The results show that HPCA outperforms PMMTM [26] and RLCM [27] in detection error. For the reason that HPCA emphasizes some important features that reflect the information of the image texture structure and these features differ greatly between cover and stego images. In this case, these important features are useful for the training process of the classifier and thus generating the steganalyzer with high detection accuracy. As a

TABLE I
PERFORMANCE COMPARISONS OF THE DIFFERENT STEGANALYZERS IN TERMS OF THE AVERAGE DETECTION ERROR ($\overline{P_E}$)

Characterization methods	Steganographic schemes	bpp 0.0044	bpp 0.0121	bpp 0.0243
HPCA	FDMS [13]	**27.19**	**17.37**	**7.97**
	PDMS [14]	**41.91**	**30.93**	**20.43**
PMMTM [26]	FDMS [13]	35.28	27.40	16.93
	PDMS [14]	45.58	39.08	30.36
RLCM [27]	FDMS [13]	38.37	26.73	15.83
	PDMS [14]	47.28	40.86	32.83

countermeasure, when steganography observes these important features change obviously after pixel toggled, the cost of pixel modifications should be set high. As a result, the important features are strongly stressed in the embedding procedure of steganography and consequently reducing their effect on steganalysis.

To achieve this goal, the weight for evaluating pixel toggling of the important features can be set large when designing the embedding distortion model. When pixel toggling causes significant changes in these features, the distortion model will set these pixel distortions large. Therefore, it avoids toggling these pixels while hiding secret information. Motivated by the L2-regularized L2-loss support vector machine (SVM) [28], a feature space among cover images and stego images is designed. Specifically, the loss of the hyperplane vector is defined as follows:

$$\min_{\mathbf{w}} \left\{ \frac{1}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w} + C\sum_{n=1}^{N}\Big[\big(\max\big(0, 1+\mathbf{w}^{\mathrm{T}}\hat{\mathbf{v}}_n^c\big)\big)^2 + \big(\max\big(0, 1-\mathbf{w}^{\mathrm{T}}\hat{\mathbf{v}}_n^s\big)\big)^2\Big]\right\} \tag{7}$$

where $\hat{\mathbf{v}}_n^c = (\mathbf{v}_n^c - \mathbf{m})\mathbf{P}$ and $\hat{\mathbf{v}}_n^s = (\mathbf{v}_n^s - \mathbf{m})\mathbf{P}$ are the projected feature vectors of cover images and stego images, respectively. We can obtain the normal vector of a separating hyperplane $\mathbf{w}$ which determines the cost weights of each feature in the feature space. $C > 0$ is a penalty factor. By adjusting the parameter C, the penalty degrees of the misclassification error are different. Because the hyperplane is trained by cover images and the corresponding stego image, there are two loss items. Specifically, we define that the cover images as negative samples with label $y = -1$, while stego images are positive samples with label $y = 1$. Therefore, $\max(0, 1+\mathbf{w}^{\mathrm{T}}\hat{\mathbf{v}}_n^c)$ is the loss of the cover image while $\max(0, 1 - \mathbf{w}^{\mathrm{T}}\hat{\mathbf{v}}_n^s)$ is the loss of the stego image. The hyperplane is trained to be the best hyperplane to discriminate cover and stego images so the loss is minimized. The five-fold cross-validation is conducted to select the best parameter C for each learning method, with the purpose of obtaining the optimal separating hyperplane.

Based on the above-mentioned discussion, the distortion model is defined by the difference between the feature vector of cover and stego images in the feature space. Given a cover image $\mathbf{X}$ and a stego image $\mathbf{Y}$, the low-dimensional feature vectors of a cover image $\hat{\mathbf{v}}^X$ and a stego image $\hat{\mathbf{v}}^Y$ are computed by

$$\begin{aligned}\hat{\mathbf{v}}^X &= \Big(\mathbf{v}^{\mathbf{X}} - \mathbf{m}\Big)\mathbf{P}\\ \hat{\mathbf{v}}^Y &= \Big(\mathbf{v}^{\mathbf{Y}} - \mathbf{m}\Big)\mathbf{P}\end{aligned} \tag{8}$$

where $\mathbf{v}^{\mathbf{X}}$ and $\mathbf{v}^{\mathbf{Y}}$ are the normalized histogram of $\mathbf{X}$ and $\mathbf{Y}$ computed using (3). Then, the distortion function $D(\mathbf{X}, \mathbf{Y})$ is written as

$$D(\mathbf{X}, \mathbf{Y}) = \left|\frac{\hat{\mathbf{v}}^X w}{\|\mathbf{w}\|} - \frac{\hat{\mathbf{v}}^Y w}{\|\mathbf{w}\|}\right| = \frac{\big|\big(\hat{\mathbf{v}}^X - \hat{\mathbf{v}}^Y\big)\mathbf{w}\big|}{\|\mathbf{w}\|} \tag{9}$$

where $\|\bullet\|$ is the modulo operation and $\mathbf{w}$ is obtained by solving (7). In fact, by projecting the image $\mathbf{X}$ onto the normal vector $\mathbf{w}$ in the feature space, (9) measures the relative distances between $\hat{\mathbf{v}}^X$ and the separating hyperplane, and so is the image $\mathbf{Y}$. The absolute difference between the two distances is finally defined as $D(\mathbf{X}, \mathbf{Y})$. It indicates that the larger the value of $D(\mathbf{X}, \mathbf{Y})$, the heavier the distortion.

When a stego image $\mathbf{Y}^{i,j}$ is obtained by only toggling the (i, j)th pixel of the cover image $\mathbf{X}$, the distortion function changes to

$$D_{i,j} \triangleq D\big(\mathbf{X}, \mathbf{Y}^{i,j}\big) = \frac{\left|\left(\hat{\mathbf{v}}^X - \hat{\mathbf{v}}^{\mathbf{Y}^{i,j}}\right)\mathbf{w}\right|}{\|\mathbf{w}\|} \tag{10}$$

where $\hat{\mathbf{v}}^{\mathbf{Y}^{i,j}}$ denotes the low-dimensional feature vector characterizing the stego image $\mathbf{Y}^{i,j}$. $D_{i,j}$ represents the pixel distortions of only toggling the (i, j)th pixel in $\mathbf{X}$. This function can be applied with the STC encoder [18] to minimize the total distortion.

B. Combined Model Based on Hybrid Feature Space

Upon the discussions in Section II-A, (9) indicates that a distortion model is associated with the feature vectors of cover and simulated stego images about HPCA. The Hotelling transformation [25] in (6) carries out different orthogonal transformation according to different sets of the simulated stego images. Consequently, it alters the feature vectors about HPCA and generates different feature spaces. In this section, we research the performance of the distortion model based on different feature spaces and propose a combined model based on a hybrid feature space.

The generation of the simulated stego images is a key element to construct feature spaces. Given an $n_1 \times n_2$ size cover image $\mathbf{X}$ and an embedding change rate ρ, the simulator $S(\mathbf{X}, \rho)$ is described as follows.

1) A pseudorandom number $\varrho_{i,j}$, $\varrho_{i,j} \in [0, 1]$ is obtained.
2) If $\varrho_{i,j} < \rho$, the pixel located at (i, j) in the image $\mathbf{X}$ is toggled.
3) Repeat steps 1 and 2 until all the pixels are traversed and then output the modified image.

It should be noted that ρ refers to not only the embedding change rate of an image but also the toggling probability of every pixel. By adjusting the embedding change rate ρ, we use the simulator $S(\mathbf{X}, \rho)$ to obtain several sets of the simulated stego images. Based on a specific set of cover and stego

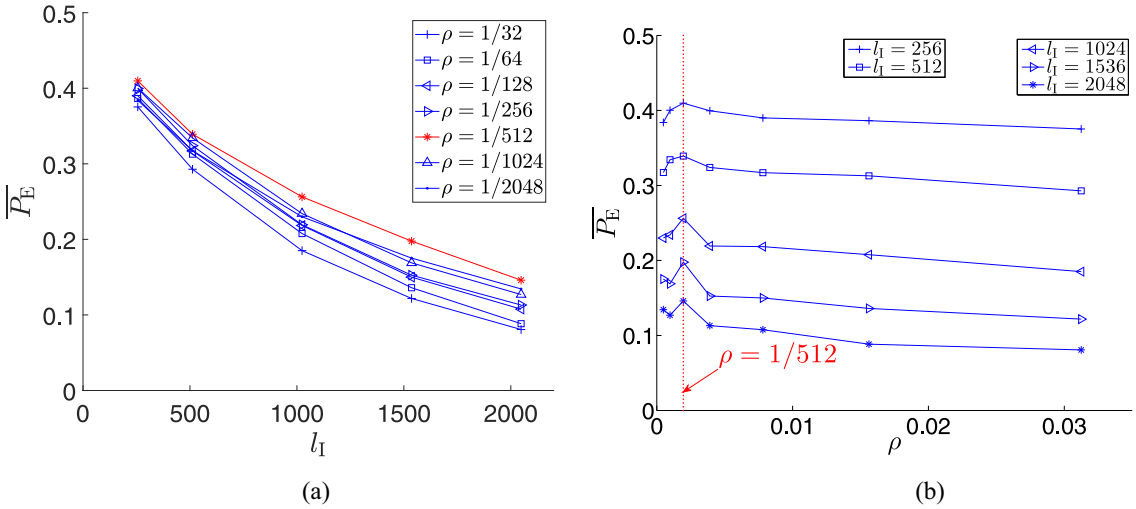


Fig. 2. Performance comparisons of the steganographic schemes applied with different distortion models D^{ρ} in terms of the average detection error ($\overline{P_E}$). The depiction of (a) and (b) is the same, but the variables on the x-axis are different. (a) Embedding message length l_{I}. (b) Embedding change rate ρ.

images, we acquire a basic embedding distortion model D^{ρ} according to (9).

In order to observe the evaluation effect of different D^{ρ}, the performance comparisons between several D^{ρ}, $\rho = 1/2^k$ and $k \in \{5, 6, \ldots, 11\}$ are set. For a fair comparison, an ideal encoder $E_I(\mathbf{X}, l_{\mathrm{I}}, D_T)$ where $D_T = D^{\rho}$ elaborated in Section V-B is employed to generate the ideal stego images. The steganalysis performance of the detectors which are trained by using PMMTM [26] with soft-margin SVMs [29] is shown in Fig. 2. The depiction of Fig. 2(a) and (b) is the same, but the variables on the x-axis are different, Fig. 2(a) is the embedding message length l_{I}, and Fig. 2(b) is the embedding change rate ρ. Fig. 2 illustrates that the steganographic scheme applied with the distortion model $D^{1/512}$ outperforms those of the other models. The simulated stego images with appropriate ρ make the distortion evaluation better and thus the ideal stego images are not easily detected by steganalyzers. In addition, Fig. 2(b) illustrates that when the payload is fixed, as ρ increases, the detection error first rises and then falls. It further indicates that the simulated stego images with too large or small ρ are unsuitable for distortion model construction. For the reason that the appropriate ρ can motivate the encoder to generate the simulated stego images whose number of toggled pixels is suitable for PCA to extract key features. From the figure, it can be concluded that the most suitable ρ is $1/512$ to generate the ideal stego images when resisting the detector with PMMTM. In practice, the designed steganographic scheme should resist various steganalysis instead of only PMMTM. Therefore, this article produces a combined distortion model below.

Based on the above discussion, to produce a combined embedding distortion model, the maximal distortion among different D^{ρ} is set as the final model

$$\max_{\rho} \left\{ D^{\rho}_{i,j} \right\}, \quad \rho \in \{1/256, 1/512, 1/1024\} \tag{11}$$

where $D^{\rho}_{i,j}$ denotes the (i, j)th pixel distortions in the basic distortion model D^{ρ} according to (10). The proposed scheme selects three basic models centered with $D^{1/512}$ which achieves the best performance in the comparisons. The maximum operation insures the steganographic algorithm to stress the maximum distortion that occurs in the three basic models.

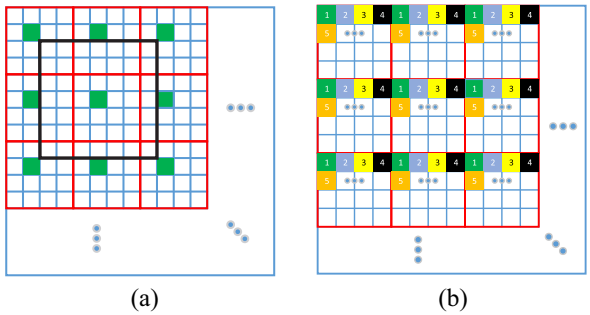


Fig. 3. Image division of the layer embedding strategy. (a) Red blocks represent the nonoverlapped blocks, green pixels are the pixels in the single layer and the region of the black block is the mutually impact area when the central green pixel is toggled. (b) Number in the pixels stands for the layer number and the layers will be embedded secret messages orderly. When embedding secret messages that are assigned to the layer, pixels in the corresponding layer will be extracted to calculate the distortion and embedded secret messages.

III. Layer Embedding

In this section, a layer embedding strategy that includes single-layer and multilayer embedding is proposed to reduce MIEM. The single-layer embedding is the basis of the multilayer embedding and the latter strategy solves the shortage of the former strategy.

A. Single-Layer Embedding

Some previous researches demonstrated that both additive distortion functions and additive-approximate distortion functions cannot capture the fact that executing the embedding modifications in a group of adjacent pixels will likely have a smaller statistical impact than changing the same number of isolated pixels [17]. As a result, the MIEMs increase and the performance of statistical undetectability decreases in practice.

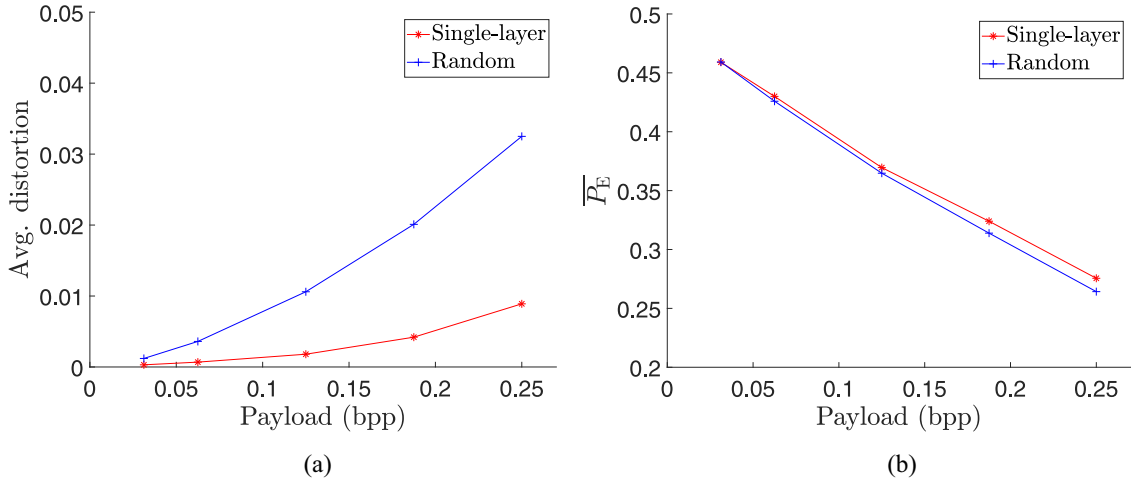


Fig. 4. Comparisons between the single-layer embedding strategy and the random embedding strategy, in terms of (a) average distortion and (b) average detection error ($\overline{P_E}$).

To reduce MIEM and improve the statistical undetectability, a layer-extraction method is proposed to construct a single layer with some pixels that do not have a mutual impact based on the proposed distortion model. In this method, an image is first divided into 4×4 nonoverlapped blocks. Then, the pixels that have the same position in every block are extracted to construct the subimage in a single layer. In the same way, the pixel distortions are also extracted to construct the subdistortion map. Fig. 3 illustrates the image division of the layer embedding strategy. Fig. 3(a) shows that the red blocks represent the 4×4 nonoverlapped blocks. The green pixels are extracted to construct the subimage. According to $D_{i,j}$, the black block illustrates the 7×7 region of mutually impact when the central green pixel is toggled. Fig. 3 also shows that the distance of any two green pixels is greater than 3 and the region of any green pixels' mutually impact is 7×7. Therefore, the image should be divided into 4×4 nonoverlapped blocks and the pixels of the subimage in the same layer does not have an impact on each other when executing the embedding modifications.

We formulate the layer embedding strategy below. There is an intermediate image denoted as $\mathbf{X}'$ that is updated after each layer embedding is finished. The subimage is used to formulate the layer embedding strategy and we have also improved it as below. If we extract the pixels and their distortions located at (a, b) of every block, we obtain the subimage $\mathbf{X}^L$ and the corresponding subdistortion map $\mathbf{D}^L$ by

$$\mathbf{X}^L_{k,l} = \mathbf{X}'_{i,j} \tag{12}$$

where $\mathbf{X}^L_{k,l}$ denotes the (k, l)th values in the intermediate image $\mathbf{X}'$. $\mathbf{D}^L$ denotes the distortion values corresponding to $\mathbf{X}^L$. The projection between (i, j) and (k, l) is defined as

$$\begin{aligned} i &= 4k + a \\ j &= 4l + b \end{aligned} \tag{13}$$

where $k \in \{0, 1, \ldots, \lfloor (n_1/4) \rfloor - 1\}$, $l \in \{0, 1, \ldots, \lfloor (n_2/4) \rfloor - 1\}$.

To verify whether the single-layer embedding strategy has taken MIEM into account and generates better stego images or not, the comparisons between the single-layer embedding strategy and random embedding strategy are set. In the random embedding strategy, the pixels in images are randomly selected but the total number of the pixels is the same as those in the single-layer embedding strategy, which ensures the same size of STC's carrier. Because of the randomness, the mutual impact cannot be avoided effectively. Then, we calculate the selected pixel distortions using the same distortion model and apply the STC encoder to embed the secret message for both two strategies. Fig. 4 shows the results of the comparisons in terms of the average distortion and average detection error. Fig. 4(a) illustrates that the single-layer embedding strategy outperforms the random embedding strategy in the average distortion. With the increase of the payload, the difference between the average distortion of the two strategies becomes much larger. Although the optimization target of STC is to minimize $D(\mathbf{X}, \mathbf{Y})$, it cannot capture the mutual impact when toggling adjacent pixels. However, there is no such effect in single-layer embedding so that the stego images have lower distortion as we expected. Fig. 4(b) shows that the single-layer embedding method also outperforms the random embedding strategy in statistical undetectability. It also shows that the higher average distortion results in lower steganalysis performance. Combining those two results, it can be concluded that the single-layer embedding strategy can improve statistical undetectability by eliminating MIEM.

B. Multilayer Embedding

Single-layer embedding makes sure that the pixels in the layer do not have a mutual impact when toggled, but it abandons too many pixels and decreases the capacity of secret messages. However, embedding secret messages directly in the cover image causes MIEM and decreases undetectability performance. Therefore, we make a balance and expand the single-layer embedding to multilayer embedding, taking MIEM into account and enlarging the secret message capacity.

Images can also be divided into multiple layers. Specifically, the single-layer embedding just makes use of one position

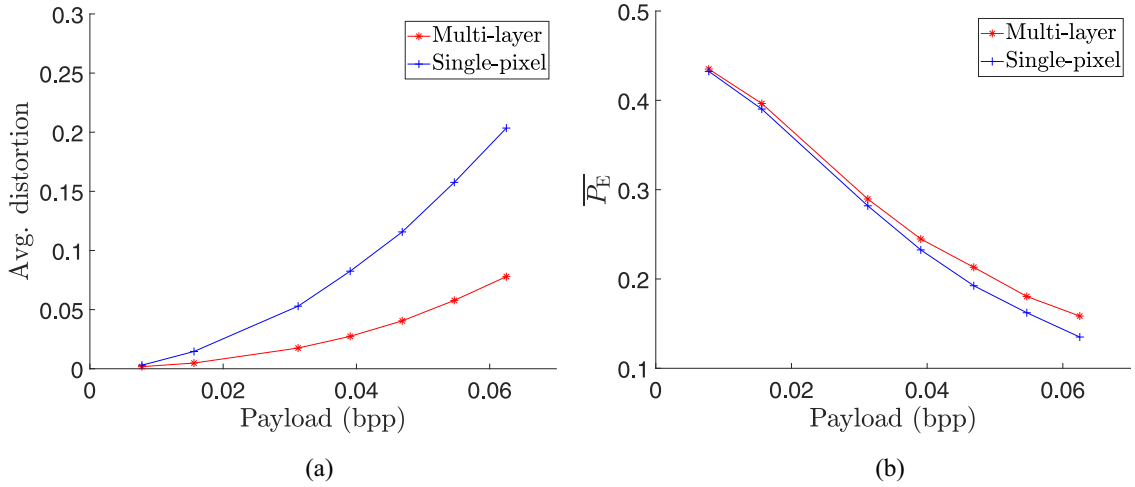


Fig. 5. Comparisons between the single-pixel embedding strategy and the multilayer embedding strategy, in terms of (a) average distortion and (b) average detection error ($\overline{P_E}$).

(a, b) of the blocks while the multilayer embedding uses all positions of the blocks. For the position (a, b), the pixels at (a, b) in every block are extracted to construct the subimage in the layer L. The relationship among L, a, and b can be represented as

$$L = 4a + b \tag{14}$$

where $a, b \in \{0, 1, 2, 3\}$. For every L, the subimage $\mathbf{sX}^L$ is generated using (12). We also need to divide the secret message **msg** into many parts. The secret message $\mathbf{msg}^L$ embedded in layer L is represented as

$$\mathbf{msg}_k^L = \mathbf{msg}_{pL+k} \tag{15}$$

where $\mathbf{msg}_k^L$ is the kth bit of $\mathbf{msg}^L$, $\mathbf{msg}_{pL+k}$ is the $(pL+k)$th bit of the original secret message, and $p = \lfloor (l_m/16) \rfloor$ ensures the secret message divided in an average way. Thus, the set of index k is as follows:

$$\begin{cases} k \in \{0, 1, \ldots, p-1\}, & \text{if } L < 15 \\ k \in \{0, 1, \ldots, l_m - pL - 1\}, & \text{if } L = 15 \end{cases} \tag{16}$$

where l_m is the length of the secret message **msg**. According to (16), the length of $\mathbf{msg}^L$ is $l_m^L = p$ for the previous 15 layers and $l_m^L = l_m - pL$ for the last layer.

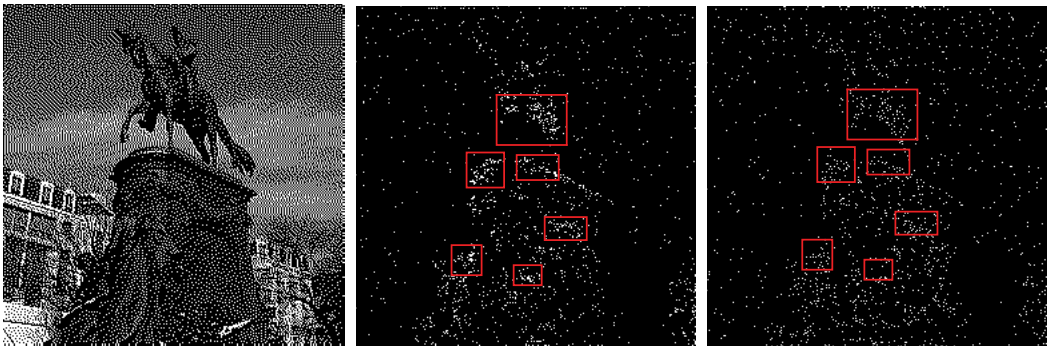
Herein, we give a brief introduction to the entire embedding procedure. As Fig. 3(b) is shown, multilayer embedding first divides the cover image into 4×4 blocks and the pixels that have the same position in their block are in the same layer. Before embedding secret messages, the secret messages are divided averagely so that each layer will be embedded the same payload of secret messages. When performing multilayer embedding, each layer is orderly calculated their distortion and embed the secret messages assigned to the layer. It is worth mentioning that after embedding one layer, the corresponding layer is updated by the embedded result and prepared to embed the next layer until all the layers are embedded with secret messages.

To verify the effectiveness of the multilayer embedding strategy, we compare the multilayer embedding strategy with the single-pixel strategy proposed in [14] that directly embeds secret messages in the entire image ignoring the order of pixels. Fig. 5(a) shows that the multilayer embedding strategy outperforms the single-pixel strategy in the average distortion. As mentioned earlier, the single-pixel strategy does not consider MIEM and it causes larger distortions. The multilayer embedding strategy updates the pixel distortions in the subsequent layer after the embedding procedure of the previous layer is executed. In the embedding procedure of each layer, the STC encoder makes $D(\mathbf{X}, \mathbf{Y})$ as the optimization target, optimizes the toggled positions, and achieves a much lower average distortion. Therefore, it provides higher steganography performance. Fig. 5(b) also shows that it outperforms the single-pixel strategy in statistical undetectability.

We conduct experiments to verify that the multilayer embedding strategy reduces MIEMs in a group of adjacent pixels and design notation G. The MIEM effect can increase the number of consecutive toggled pixels, which is large in the single-pixel strategy, while the multilayer embedding strategy avoids this phenomenon well and thus reduces MIEM. Fig. 6 shows one of the embedding modification examples of the two strategies. To quantify such an effect, we design a method to evaluate the effect of toggling adjacent pixels. When a pixel is toggled, its adjacent pixels are defined as the pixels in the 7×7 block centered with it. As is mentioned, the 7×7 block is the region of mutually impact when the central pixel is toggled. The adjacent toggled pixels pair is defined as the pair of the central pixel of the block and other toggled pixels in the same block. More adjacent toggled pixel pairs demonstrate that more pixels that have MIEM each other are toggled simultaneously. This phenomenon is not considered by the single-pixel strategy and decreases statistical security. The group of adjacent toggled pixel pairs G can be represented as

$$\begin{aligned} G = \big\{ (i, j, p, q) | \Delta_{i,j} = 1, \Delta_{p,q} = 1 \\ |i - p| \le 3, |j - q| \le 3 \\ i \ne p \quad \text{or} \quad j \ne q \big\} \end{aligned} \tag{17}$$

where $\Delta = |\mathbf{X} - \mathbf{Y}|$ denotes the embedding changes between cover image $\mathbf{X}$ and the stego image $\mathbf{Y}$, and thus $\Delta_{i,j}$ is 1 when the (i, j)th pixel is toggled, otherwise is 0, and so is $\Delta_{p,q}$.



(a) (b) (c)

Fig. 6. Embedding modifications of the single-pixel strategy and the multilayer embedding strategy when the payload is 0.0313 bpp. (a) Cover image. The embedding modifications of (b) single-pixel strategy and (c) multilayer embedding strategy.

Then, the number of adjacent toggled pixel pairs $N(\mathbf{X}, \mathbf{Y})$ can be represented as

$$N(\mathbf{X}, \mathbf{Y}) = \frac{1}{2}|G| \tag{18}$$

where $|G|$ denotes the size of the group G and we take it a half considering the symmetric structure.

Fig. 7 shows the result of average $N(\mathbf{X}, \mathbf{Y})$ between the two strategies. It can be found that in the multilayer embedding strategy, the number of toggled pixel pairs that have a mutual effect is smaller compared with the other strategy. $N(\mathbf{X}, \mathbf{Y})$ have the same trend as average distortion and the steganalysis results. We consider that there are two main reasons. On the one hand, we use the single-layer embedding strategy to make sure the pixels in the layer do not have the mutual effect to make the best use of the STC encoder. On the other hand, we embed secret messages layer by layer using the previous layers as a prior condition and taking previous modifications into account. Combing those benefits, the multilayer embedding progresses much compared with the single-pixel strategy. With the increase of the payload, the difference between the two strategies becomes much larger. That is because the higher payload of the secret messages, the more pixels are toggled, causing much larger MIEM, which is also explained by the $N(\mathbf{X}, \mathbf{Y})$ experiment.

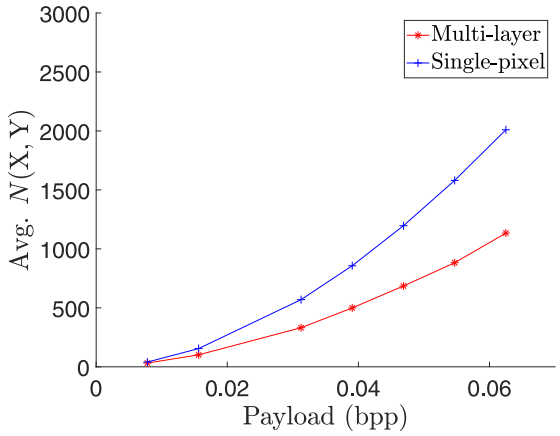


Fig. 7. Comparison between the single-pixel strategy and the multilayer embedding strategy in average $N(\mathbf{X}, \mathbf{Y})$.

IV. Proposed Steganographic Scheme

The entire steganographic scheme is presented detailedly in this section, including the embedding and extraction procedure.

A. Embedding Procedure

The block diagram of the embedding procedure is shown in Fig. 8. Given a cover image $\mathbf{X}$ and a secret binary message $\mathbf{msg}$, the embedding procedure $E_{\mathrm{STC}}(\mathbf{X}, \mathbf{msg})$ consists of the following steps.

1) A secret message $\mathbf{msg}$ is averagely divided into 16 nonoverlapped segments using (15), successively denoted as $\mathbf{msg}^L$ whose length is l_m^L.
2) Initialize the intermediate image $\mathbf{X}' = \mathbf{X}$ and $L = 0$ in the beginning of the procedure.
3) Based on the value of L, the distortion map $\mathbf{D}^L$ corresponded to the image $\mathbf{X}^L$ is extracted and computed using the intermediate image $\mathbf{X}'$.
4) Reshape $\mathbf{X}^L$ and $\mathbf{D}^L$ into a 1-D vector and scramble them using a seed s^L, denoted as $\mathbf{uD}^L$ and $\mathbf{uX}^L$, respectively.
5) Apply $\mathbf{uD}^L$, $\mathbf{uX}^L$, and $\mathbf{msg}^L$ with the STC encoder. The embedded pixels $\mathbf{u\ddot{X}}^L$ are obtained by STC encoding.
6) Descramble $\mathbf{u\ddot{X}}^L$ using the seed s^L and reshape it into the size of subimage $\mathbf{X}^L$, denote as $\ddot{\mathbf{X}}^L$.
7) Update the intermediate image $\mathbf{X}'$ by using $\ddot{\mathbf{X}}^L$ to replace $\mathbf{X}^L$ originally extracted from the image $\mathbf{X}^L$.
8) If $L < 16$, repeat steps 3–7, otherwise, the stego image $\mathbf{Y}$ is obtained, that is, $\mathbf{Y} = \mathbf{X}'$.

Note that both the receiver and sender should know the random seeds s^L, which are used to scramble. They can prepare the parameter of seeds in advance.

B. Extraction Procedure

The block diagram of the extraction procedure is shown in Fig. 9. Given a stego image $\mathbf{Y}$, the length of the original messages l_m and the random seed s^L. We can elaborate the extraction procedure $D_{\mathrm{STC}}(\mathbf{Y}, l_m)$ as the following steps.

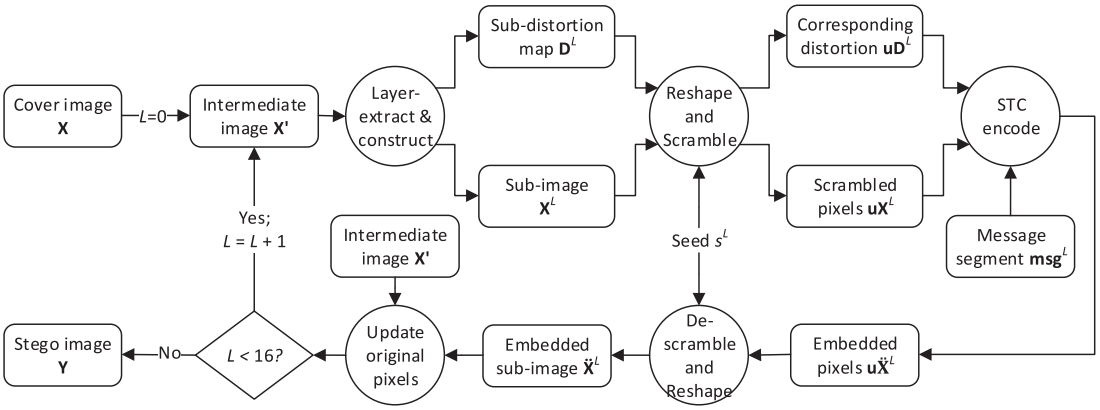


Fig. 8. Embedding block diagram.

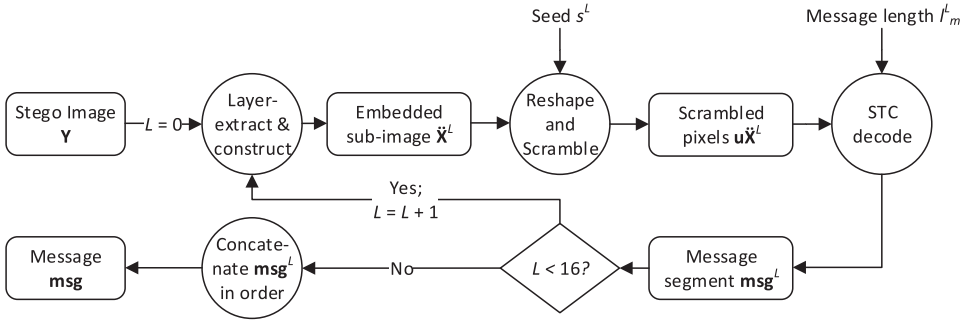


Fig. 9. Extraction block diagram.

1) Based on the value of L, a stego image $\mathbf{Y}$ is correspondingly extracted to construct the embedded subimage $\ddot{\mathbf{X}}^L$ and l_m^L is obtained according to (16).
2) Reshape $\ddot{\mathbf{X}}_L$ into a 1-D vector and scramble it using a seed s^L that the receiver and sender share, denoted as $\mathbf{u}\ddot{\mathbf{X}}^L$.
3) Apply message length l_m^L and $\mathbf{uX}^L$ with the STC decoder, and then obtain message segment $\mathbf{msg}^L$.
4) If $L < 16$, repeat steps 1–3, otherwise, the message $\mathbf{msg}$ is obtained by successively concatenate $\mathbf{msg}^L$ in order.

V. Experimental Results

A. Experimental Setup

All the experiments are conducted on the BOSSbase ver.1.01 database [30], which contains 10 000 grayscale images of size 512×512 pixels. In the experiments, we resample all the grayscale images into the size of 256×256 pixels and transfer them into halftone images using error diffusion [31]. It is worth mentioning that error diffusion is a classic and effective method that guarantees the visual quality of the generated halftone images well. The proposed scheme does not depend on a specific halftoning technique according to the statistical distortion model. In this article, all experiments are conducted on the image datasets that are generated based on the error diffusion method. Then, the image database is divided into two sets of equal size, one used for training and the other used for accuracy evaluation. The performance is measured by the detection error rate P_E under the equal probability of cover and stego images, defined as

$$P_E = \frac{1}{2}(P_{\mathrm{FA}} + P_{\mathrm{MD}}) \tag{19}$$

where P_{FA} and P_{MD} stand for the probabilities of false alarm and miss detection, respectively. The final statistical security is accessed by the average error rate $\overline{P_E}$, which is calculated over ten random splits of the datasets. It should be noted that the statistical security of schemes is stronger when $\overline{P_E}$ is larger.

B. Comparisons of the Distortion Model

Experiments are conducted to compare the steganalysis performance among different distortion models [13], [14]. Note that the proposed model used in the experiments is the combined model that has been elaborated in Section II-B and the dimensionality of the histogram is reduced to 500 experimentally. A good distortion model should effectively evaluate the embedding distortion and it can improve the statistical security of steganography [4], [20]. In this section, to evaluate the performance of the proposed distortion model, we compare it with FDM [13] and PDM [14], which have been discussed in Section I.

For the fair comparisons, an ideal encoder $E_I(\mathbf{X}, l_{\mathrm{I}}, D_T)$ is designed to simulate an optimal modification, which includes the following steps.

1) Use the test distortion model D_T to calculate a distortion map that corresponds to the image $\mathbf{X}$.

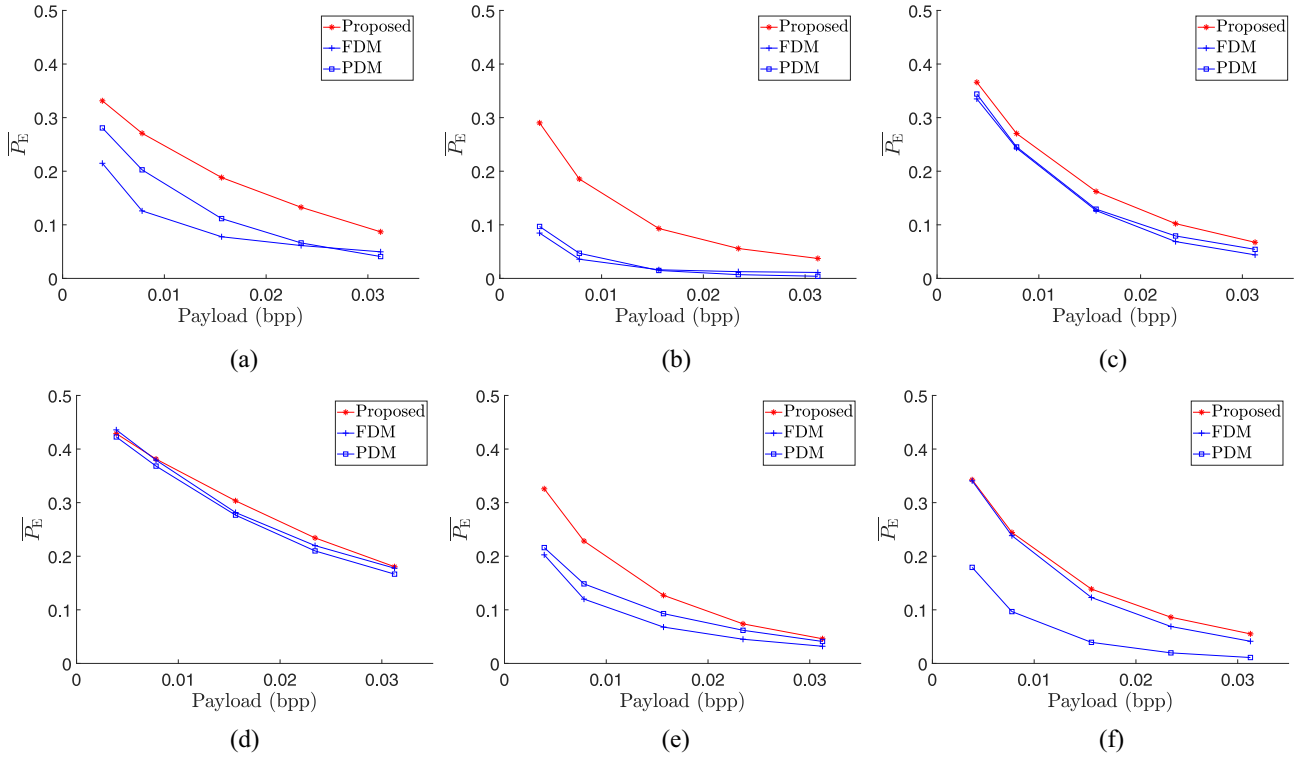


Fig. 10. Performance comparisons of the steganographic schemes with different distortion models D^{ρ} in terms of the average detection error ($\overline{P_E}$), while attacked by the detectors applied with (a) PMMTM [26], (b) LGLTP [22], (c) RLCM [27], (d) RLGL [32], (e) DLCM [33], and (f) PHD [34].

2) Obtain the embedding positions which causes the lowest distortion and replacing the pixel in the selected position with a pseudorandom binary bit.
3) Set pixels within 3×3 block that surround the selected pixel in step 2 to be nonoptional.
4) Repeat steps 2 and 3 until l_{I} length pixels in $\mathbf{X}$ are replaced and output the modified image.

Based on the encoder, the stego images with the lowest total distortion are correspondingly obtained by a specific distortion model. When l_{I} parameter changes, a better distortion model will generate the stego images with better statistical security. Via the designed encoder, we compare the security of different distortion models combined with $E_I(\mathbf{X}, l_I, D_T)$ in the cases of $l_{\mathrm{I}} \in \{256, 512, 1024, 1536, 2048\}$, that is, the corresponding payloads of {0.0039, 0.0078, 0.0156, 0.0234, 0.0313} bpp.

The statistical security is represented by the performance for resisting the state-of-the-art steganalysis [22], [26], [27], [32]. PMMTM [26] was proposed by focusing on the relationship between pixel mesh transitions. LGLTP [22] is presented to consider a larger local texture pattern that contains more information on the image content edge. RLCM [27] is the gray-level run length combined with co-occurrence matrices and RLGL [32] is the gray-level run length combined with gap length matrices with the purpose of capturing the different gray runs and the interpixel relationship between the cover and stego images. DLCM [33] is extracted by a distortion-level co-occurrence matrix that represents the toggled pixels causing significant changes in neighboring distortion. PHD [34] was proposed by Chiew and Pieprzyk to capture the histogram difference between cover and stego images based on 3×3 pixel blocks. PMMTM, RLCM, RLGL, DLCM, and PHD are combined with soft-margin SVMs [29] with an optimized Gaussian kernel to construct the detectors while LGLTP combined with the ensemble classifiers (ECs) [35] with the Fisher linear discriminant as the base learner are trained to build the detector. The detection performance is measured by $\overline{P_E}$ which has been mentioned in Section V-A.

Fig. 10 illustrates that the proposed distortion model outperforms the others, which indicates that the proposed model evaluates the embedding distortion more effectively. It is observed that the scheme with the proposed distortion model obtains a significant improvement while attacked by the detectors applied with PMMTM and LGLTP. Stego images generated according to both FDM and PDM are easily identified by the detector applied with LGLTP shown in Fig. 10(b), but the steganographic scheme applied with the proposed distortion model achieves high steganalysis performance because the statistical region of the proposed model is larger than those of LGLTP. Fig. 10(c), (d), and (f) shows that the performance obtained by the proposed distortion model is slightly better than the other models. In brief, the proposed distortion model can better evaluate the embedding distortion compared with FDM and PDM, thus applying it with the STC encoder can improve the statistical undetectability.

C. Comparisons of Steganographic Schemes

We have evaluated the performance of the distortion model and the embedding strategy separately. In this section, we combine them and evaluate the performance of the entire

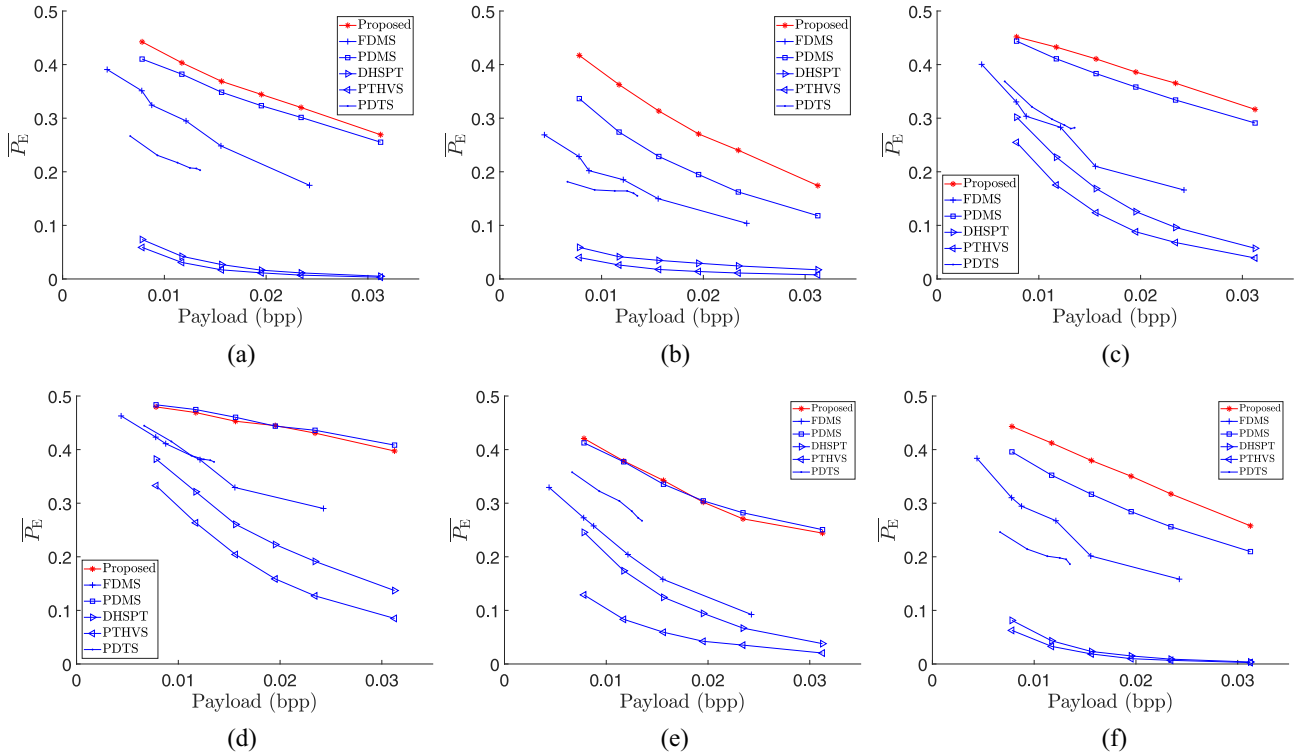


Fig. 11. Performance comparisons of different steganographic schemes in terms of the average detection error ($\overline{P_E}$), while attacked by the detectors applied with (a) PMMTM [26], (b) LGLTP [22], (c) RLCM [27], (d) RLGL [32], (e) DLCM [33], and (f) PHD [34].

schemes to compare the statistical security of the different steganographic schemes.

Schemes proposed in [13] (denoted as FDMS), [14] (denoted as PDMS), [11] (denoted as DHSPT), [12] (denoted as PTHVS), and [36] (denoted as PDTS) are employed for comparisons. Both FDMS and PDMS minimize the toggling distortion based on the STC encoder, but superpixels are regarded as STC's carriers in the former scheme while the latter scheme considers single pixels. DHSPT is presented to improve its basic scheme DHPT by wisely choosing the slave pixels so that reducing the "salt-and-pepper" clusters. On the basis of DHSPT, to further improve the selection of candidate slave pixels, PTHVS is proposed by designing a visual distortion model according to a larger local region. PDTS uses the pixel density transition to embed messages and optimizes the visual quality by selecting suitable density blocks.

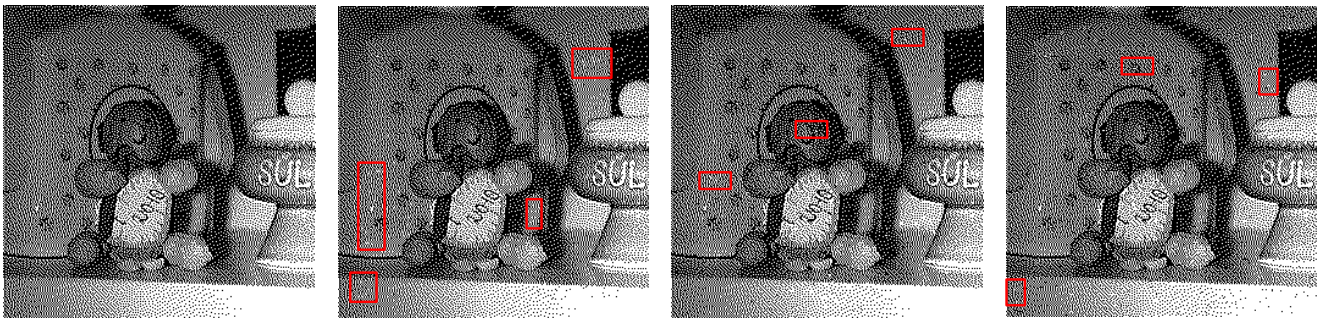
Throughout the experiments, the pseudorandom binary sequences are used as secret messages. The message length l_m is set via different parameters of different schemes. We compare the proposed scheme with other steganographic schemes under the average payloads of the image database. In FDMS, the scheme set the length of message segment θ_m as {8, 16}, the number of elements in superpixels $\theta_{\mathfrak{I}}$ as $\{3^2, 4^2, 5^2\}$, and the length of cover vectors as 8^2, that is, the payloads of {0.0044, 0.0078, 0.0088, 0.0121, 0.0156, 0.0243} bpp. The other steganographic schemes are only limited by the parameter l_m that is set as {512, 768, 1024, 1280, 1536, 2048}, that is, the payloads of {0.0078, 0.0117, 0.0156, 0.0195, 0.0234, 0.0313} bpp. The steganalysis and the detection performance criterion used in Section V-B are still employed here.

The performance comparisons of different steganographic schemes are shown in Fig. 11. It can be observed that DHSPT and PTHVS schemes do not obtain high statistical security when resisting all the steganalysis presented in the figure, for the reason that the number of embedding modifications in their schemes is large which is equal to the secret message length and thus causing a large embedding distortion. The other schemes use the STC encoder to embed the secret messages with the same length by toggling fewer pixels, which results in less total distortion. Compared with FDMS and PDMS, the proposed scheme obtains the best security while attacked by the detectors applied with PMMTM, LGLTP, RLCM, and PHD. Particularly attacked by the detector applied with LGLTP and PHD, the proposed scheme achieves a significant improvement compared with PDMS and DLCM which has the second-best security. Fig. 11(d) and (e) illustrates that the proposed scheme and PDMS can provide equal security performance but in other steganalysis schemes, the proposed method outperforms PDMS. In conclusion, the proposed method outperforms other state-of-the-art methods under multiple steganalysis methods' attack and achieves the highest security performance.

D. Comparisons of Visual Quality

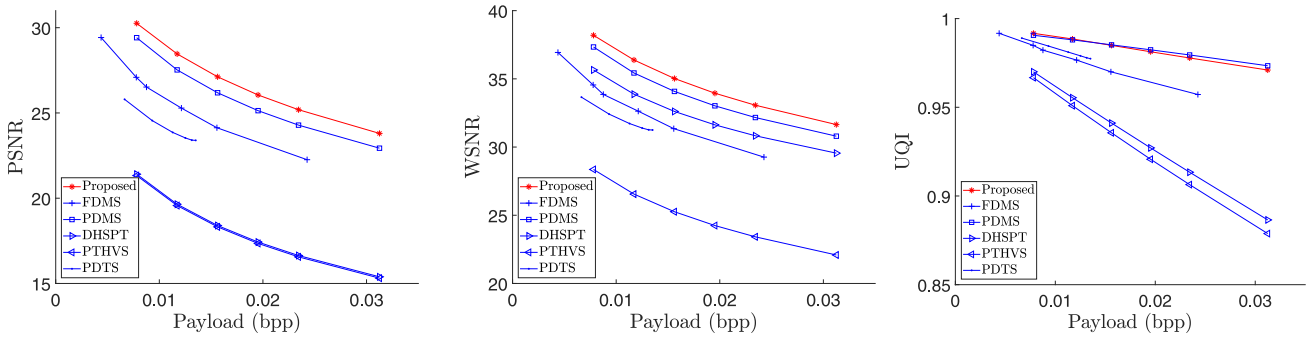
To better evaluate the visual imperceptibility of the stego images in the compared steganographic schemes, we have supplemented human visual perception and objective vision imperceptibility as follows.

For human visual perception, we take one of the complex texture halftone images as an example and show the unnatural



(a) (b) (c) (d)

Fig. 12. Image visual quality comparison of different steganography schemes on halftone image. (a) Original halftone image of size 256 × 256. (b)–(d) Stego images with 0.023 bpp secret messages embedded by FDMS [13], PDMS [14], and the proposed scheme and the red boxes of each subfigure show the unnatural changes after embedding the same secret messages.



(a) (b) (c)

Fig. 13. Visual quality comparisons of different steganographic schemes. (a) PSNR. (b) WSNR [37]. (c) UQI [38].

changes after embedding the same secret messages. Fig. 12 illustrates the experimental results. The red boxes of each subfigure show the unnatural changes after embedding the same secret messages. It shows that FDMS [13] makes more unnatural changes compared to PDMS [14] and the proposed scheme. However, it is hard to directly compare PDMS and the proposed scheme. Therefore, we also conduct experiments on the objective visual quality to compare the visual quality more precisely.

For objective vision imperceptibility, we have conducted experiments on different objective visual quality measurements, peak signal-to-noise ratio (PSNR), weighted signal-to-noise ratio (WSNR) [37], and universal image quality index (UQI) [38]. PSNR in halftone image steganography measures the number of toggled pixels. A higher PSNR means better visual imperceptibility and the stego image is perceived more similar to the cover image. WSNR [37] evaluates the visual performance for halftone images by weighting the stego image according to a contrast sensitivity function (CSF). CSF is the linear approximation of the human visual system assuming that the human eyes do not focus on one point but freely moves the eyes around the image. A higher WSNR means higher visual quality. UQI [38] is designed for halftone images, which considers the different attributes, including brightness, contrast, texture, orientation, etc. The dynamic range of UQI is $[-1, 1]$, and the best value 1 is achieved only if the stego image is fully equal to the cover image.

Fig. 13 shows that in terms of PSNR and WSNR, the stego images generated by the proposed scheme compared with the other schemes maintain the visual quality better, for the reason that the proposed scheme can toggle fewer pixels at the same embedding capacity. In terms of UQI, the performance of the proposed scheme is close to those of PDMS [14]. PDMS presented a PDM that aims at predicting the pixel value statistically so that the stego image is as close as possible to the cover image. In conclusion, the objective visual quality of stego images generated by the proposed scheme achieves good performance.

E. Analysis of Computational Complexity

The complexity of the proposed method can be analyzed in two parts. The first part is the calculation of pixel distortion. Since the hyperplane vector is pretrained and can be directly used in distortion measurement, the complexity of the calculation of pixel distortion is approximately equal to the size of the pattern block, which is a constant 4×4. The second part is the complexity of multilayer embedding. Because for each layer, only the pixels in the layer are used to embed secret messages, the distortion of each pixel is calculated once in the multilayer embedding. Assume the size of cover image is $n_1 \times n_2$. The time complexity of embedding secret messages is $\mathcal{O}(n_1 \times n_2)$ while the extraction is the reverse process of embedding secret messages and have the same complexity.

It is worth mentioning that the proposed scheme only needs to compute the pixel distortion once for each pixel, which is equal to the single-pixel embedding strategy. Therefore, the two strategies have the same computational complexity, which is $\mathcal{O}(n_1 \times n_2)$. In addition, those schemes based on the single-pixel embedding strategy are performed by calculating the distortion of each pixel once and have the same time complexity. As a result, although the proposed scheme performs a more complex embedding strategy, it has the same computational complexity but succeeds to earn improvement in anti-steganalysis performance.

We have also made our code of the proposed scheme open-source in the GitHub.[1] Everyone is able to access the code of the proposed scheme.

VI. Conclusion

In this article, a halftone image steganographic scheme based on a feature space and layer embedding was proposed to achieve high statistical security. A better design of the image model can enhance anti-steganalysis performance. A characterization method was first designed according to the statistics of 4×4 pixel blocks to construct the feature space. Upon it, a generalized blind steganalyzer with high steganalysis performance was used as a guide for designing a distortion model. As a result, an image model was defined in the hybrid feature space to measure the embedding distortion, which outperforms some state-of-the-art models. Beyond that, the embedding strategy also has a huge impact on the statistical security of steganography. The single-layer and multilayer embedding strategies were proposed based on the proposed distortion model. In single-layer embedding, the calculated pixel distortions are independent and thus it eliminates the MIEMs. The multilayer embedding strategy is expanded from the single-layer embedding to solve the shortage of low message capacity and also took MIEM into account. In brief, the layer embedding strategy reduced MIEM and improved statistical security. Comparisons with prior schemes have demonstrated that the entire steganographic scheme can achieve high statistical security of anti-steganalysis.

References

[1] J. Wang, J. Ni, X. Zhang, and Y. Shi, "Rate and distortion optimization for reversible data hiding using multiple histogram shifting," *IEEE Trans. Cybern.*, vol. 47, no. 2, pp. 315–326, Feb. 2017.

[2] J. Guo and Y. Liu, "Hiding multitone watermarks in halftone images," *IEEE Trans. Multimedia*, vol. 17, no. 1, p. 65, Jan.–Mar. 2010.

[3] X. Cao, L. Du, X. Wei, D. Meng, and X. Guo, "High capacity reversible data hiding in encrypted images by patch-level sparse representation," *IEEE Trans. Cybern.*, vol. 46, no. 5, pp. 1132–1143, May 2016.

[4] T. Pevný, T. Filler, and P. Bas, "Using high-dimensional image models to perform highly undetectable steganography," in *Proc. Int. Workshop Inf. Hiding Lecture Notes Comput. Sci*, 2010, pp. 161–177.

[5] W. Zhang, Z. Zhang, L. Zhang, H. Li, and N. Yu, "Decomposing joint distortion for adaptive steganography," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 10, pp. 2274–2280, Oct. 2017.

[6] Y. Wang, G. Zhu, S. Kwong, and Y. Shi, "A study on the security levels of spread-spectrum embedding schemes in the WOA framework," *IEEE Trans. Cybern.*, vol. 48, no. 8, pp. 2307–2320, Aug. 2018.

[1] https://github.com/stego-researching/FeatureSpace

[7] R. Eschbach and K. T. Knox, "Error-diffusion algorithm with edge enhancement," *J. Opt. Soc. Amer. A*, vol. 8, no. 12, pp. 1844–1850, 1991.

[8] R. W. Floyd and L. Steinberg, "An adaptive algorithm for spatial grayscale," *Proc. Soc. Inf. Display*, vol. 17, no. 2, pp. 75–77, 1976.

[9] Y. Liu and J. Guo, "Dot-diffused halftoning with improved homogeneity," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4581–4591, Nov. 2015.

[10] M. Analoui and J. P. Allebach, "Model-based halftoning using direct binary search," in *Proc. Soc. Photo Opt. Instrum. Eng. (SPIE) Conf. Series*, vol. 1666, Aug. 1992, pp. 96–108.

[11] M. S. Fu and O. C. L. Au, "Data hiding watermarking for halftone images," *IEEE Trans. Image Process.*, vol. 11, no. 4, pp. 477–484, Apr. 2002.

[12] J. Guo, "Improved pair toggling data hiding by cooperating human visual system in halftone images," in *Proc. Int. Conf. Acoust.*, vol. 2. Honolulu, HI, USA, 2007, pp. 285–288.

[13] B. Feng, W. Lu, and W. Sun, "Secure binary image steganography based on minimizing the distortion on the texture," *IEEE Trans. Inf. Forensics Security*, vol. 10, pp. 243–255, 2015.

[14] Y. Yeung, W. Lu, Y. Xue, J. Huang, and Y. Shi, "Secure binary image steganography with distortion measurement based on prediction," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 5, pp. 1423–1434, May 2020.

[15] W. Lu, L. He, Y. Yeung, Y. Xue, H. Liu, and B. Feng, "Secure binary image steganography based on fused distortion measurement," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 6, pp. 1608–1618, Jun. 2019.

[16] B. Li, M. Wang, X. Li, S. Tan, and J. Huang, "A strategy of clustering modification directions in spatial image steganography," *IEEE Trans. Inf. Forensics Security*, vol. 10, pp. 1905–1917, 2015.

[17] T. Denemark and J. Fridrich, "Improving steganographic security by synchronizing the selection channel," in *Proc. 3rd Workshop Inf. Hiding Multimedia Security (IH&MMSec)*, 2015, pp. 5–14.

[18] T. Filler, J. Judas, and J. Fridrich, "Minimizing additive distortion in steganography using syndrome-trellis codes," *IEEE Trans. Inf. Forensics Security*, vol. 6, pp. 920–935, 2011.

[19] B. Li, M. Wang, J. Huang, and X. Li, "A new cost function for spatial image steganography," in *Proc. Int. Conf. Image Process.*, Paris, France, 2014, pp. 4206–4210.

[20] J. Fridrich and T. Filler, "Practical methods for minimizing embedding impact in steganography," in *Proc. Conf. Security Steganogr. Watermarking Multimedia Contents*, vol. 6505, 2007, Art. no. 650502.

[21] Z. Li and A. G. Bors, "Selection of robust and relevant features for 3-D steganalysis," *IEEE Trans. Cybern.*, vol. 50, no. 5, pp. 1989–2001, May 2020.

[22] J. Chen, W. Lu, Y. Fang, X. Liu, Y. Yeung, and Y. Xue, "Binary image steganalysis based on local texture pattern," *J. Vis. Commun. Image Represent.*, vol. 55, pp. 149–156, Aug. 2018.

[23] J. Kodovskỳ and J. Fridrich, "On completeness of feature spaces in blind steganalysis," in *Proc. ACM Workshop Multimedia Security*, 2008, pp. 123–132.

[24] T. Pevný, P. Bas, and J. Fridrich, "Steganalysis by subtractive pixel adjacency matrix," *IEEE Trans. Inf. Forensics Security*, vol. 5, no. 2, pp. 215–224, Jun. 2010.

[25] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, 4th ed. Amsterdam, The Netherlands: Elsevier Pte Ltd., 2009.

[26] B. Feng, W. Lu, and W. Sun, "Binary image steganalysis based on pixel mesh Markov transition matrix," *J. Vis. Commun. Image Represent.*, vol. 26, pp. 284–295, Jan. 2015.

[27] K. L. Chiew and J. Pieprzyk, "Binary image steganographic techniques classification based on multi-class steganalysis," in *Information Security, Practice and Experience*. Berlin, Germany: Springer, 2010, pp. 341–358.

[28] R. Fan, K. Chang, C. Hsieh, X. Wang, and C. Lin, "LIBLINEAR: A library for large linear classification," *J. Mach. Learn. Res.*, vol. 9, pp. 1871–1874, Jun. 2008.

[29] C. Chang and C. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, p. 27, 2011.

[30] P. Bas, T. Filler, and T. Pevný, "'Break our steganographic system': The ins and outs of organizing boss," in *Proc. 13th Int. Workshop Inf. Hiding Hiding Lecture Notes Comput. Sci.*, 2011, pp. 59–70.

[31] B. E. Bayer, "An optimum method for two-level rendition of continuous-tone pictures," in *Proc. IEEE Int. Commun. Conf.*, 1973, pp. 2611–2615.

[32] K. L. Chiew and J. Pieprzyk, "Blind steganalysis: A countermeasure for binary image steganography," in *Proc. Int. Conf. Availability Rel. Security*, Krakow, Poland, 2010, pp. 653–658.

[33] J. Chen *et al.,* "Binary image steganalysis based on distortion level co-occurrence matrix," *Comput. Mater. Continua*, vol. 55, no. 2, pp. 201–211, 2018.
[34] K. L. Chiew and J. Pieprzyk, "Estimating hidden message length in binary image embedded by using boundary pixels steganography," in *Proc. Int. Conf. Availability Rel. Security*, Krakow, Poland, 2010, pp. 683–688.
[35] J. Kodovskỳ, J. Fridrich, and V. Holub, "Ensemble classifiers for steganalysis of digital media," *IEEE Trans. Inf. Forensics Security*, vol. 7, pp. 432–444, 2012.
[36] W. Lu, Y. Xue, Y. Yeung, H. Liu, J. Huang, and Y. Shi, "Secure halftone image steganography based on pixel density transition," *IEEE Trans. Depend. Secure Comput.*, early access, Aug. 6, 2019, doi: 10.1109/TDSC.2019.2933621.
[37] M. Valliappan, B. L. Evans, D. A. Tompkins, and F. Kossentini, "Lossy compression of stochastic halftones with JBIG2," in *Proc. Int. Conf. Image Process. (Cat. 99CH36348)*, vol. 1. Kobe, Japan, 1999, pp. 214–218.
[38] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.

Wei Lu (Member, IEEE) received the B.S. degree in automation from Northeast University, Shenyang, China, in 2002, and the M.S. and Ph.D. degrees in computer science from Shanghai Jiao Tong University, Shanghai, China, in 2005 and 2007, respectively.

He was a Research Assistant with Hong Kong Polytechnic University, Hong Kong, from 2006 to 2007. He is currently a Professor with the School of Data and Computer Science, Sun Yat-sen University, Guangzhou, China. His research interests include multimedia forensics and security, data hiding, privacy protection, and computer vision.

Prof. Lu is an Associate Editor for *Signal Processing* and the *Journal of Visual Communication and Image Representation*.

Junjia Chen received the B.S. degree in information engineering from South China Normal University, Guangzhou, China, in 2017, and the M.S. degree in computer science from Sun Yat-sen University, Guangzhou, in 2020.

His research interests include multimedia security and data hiding.

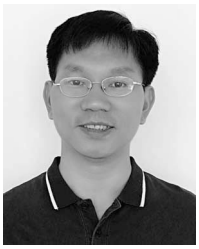
Junhong Zhang received the B.S. degree in computer science and technology from Sun Yat-sen University, Guangzhou, China, in 2018, where he is currently pursuing the M.S. degree with the School of Data and Computer Science.

His research interests include multimedia security and data hiding.

Jiwu Huang (Fellow, IEEE) received the B.S. degree from Xidian University, Xi'an, China, in 1982, the M.S. degree from Tsinghua University, Beijing, China, in 1987, and the Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences, Beijing, in 1998.

He is currently a Professor with the College of Information Engineering, Shenzhen University, Shenzhen, China. His current research interests include multimedia forensics and security.

Prof. Huang was the General Co-Chair of IEEE Workshop on Information Forensics and Security in 2013, and the TPC Co-Chair of IEEE Workshop on Information Forensics and Security in 2018. He is an Associate Editor for IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY. He is a Member of the IEEE Signal Processing Society Information Forensics and Security Technical Committee.



Jian Weng (Member, IEEE) received the B.S. and M.S. degrees in computer science and engineering from the South China University of Technology, Guangzhou, China, in 2000 and 2004, respectively, and the Ph.D. degree in computer science and engineering from Shanghai Jiao Tong University, Shanghai, China, in 2008.

From 2008 to 2010, he held a postdoctoral position with the School of Information Systems, Singapore Management University, Singapore. He is currently a Professor and the Vice President with Jinan University, Guangzhou. He has published over 100 papers in cryptography and security conferences and journals, such as CRYPTO, EUROCRYPT, ASIACRYPT, IEEE TRANSACTIONS ON CLOUD COMPUTING, *Protein Kinase C*, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, and IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING. His research interests include public-key cryptography, cloud security, and blockchain.

Prof. Weng served as the PC Co-Chair or PC Member for more than 30 international conferences. He also serves as an Associate Editor of IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY.

Yicong Zhou (Senior Member, IEEE) received the B.S. degree in electrical engineering from Hunan University, Changsha, China, in 1992, and the M.S. and Ph.D. degrees in electrical engineering from Tufts University, Medford, MA, USA, in 2008 and 2010, respectively.

He is currently an Associate Professor and the Director of the Vision and Image Processing Laboratory, Department of Computer and Information Science, University of Macau, Macau, China. His research interests include image processing, computer vision, machine learning, and multimedia security.

Dr. Zhou was a recipient of the Third Price of Macao Natural Science Award in 2014 and 2020. He is the Co-Chair of the Technical Committee on Cognitive Computing in the IEEE Systems, Man, and Cybernetics Society. He serves as an Associate Editor for IEEE TRANSACTIONS ON NEUTRAL NETWORKS AND LEARNING SYSTEMS, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, and four other journals. He is a Senior Member of the International Society for Optical Engineering.