Evolving Distortion Function by Exploiting the Differences Among Comparable Adaptive Steganography

Wenbo Zhou
CAS Key Laboratory of
Electro-magnetic Space Information
University of Science
and Technology of China
Hefei, China
Email:welbeckz@mail.ustc.edu.cn

Weiming Zhang
CAS Key Laboratory of
Electro-magnetic Space Information
University of Science
and Technology of China
Hefei, China
Tel.:0551-63600863
Email:zhangwm@ustc.edu.cn

Nenghai Yu
CAS Key Laboratory of
Electro-magnetic Space Information
University of Science
and Technology of China
Hefei, China
Tel.:0551-63600681

***Abstract*—So far, the most effective model for adaptive steganography is to minimize a well-defined distortion function, in which the distortion function determines the modification probability (MP) of each pixel. We found that the MPs of some pixels calculated by a group of steganographic methods may be very different even though these methods have close performances in resisting the detection of steganalysis. We call such pixels as controversial pixels, and consider that steganalysis is not sensitive to such pixels. Therefore we can assign more payloads to the controversial pixels by increasing MPs on them. We call this strategy as the rule of Controversial Pixels Prior (CPP). Taking the state-of-art methods {WOW, UNIWARD} and {HILL, MVG} as two pairs of examples, we show that the principle of CPP can improve the security of state of the art steganographic algorithms for spatial images.**

***Index Terms*—Steganography, evolution, distortion function, controversial pixels, modification probability, steganalysis**

I. Introduction

St eganography is a technique for covert communication, which aims to hide secret messages into ordinary digital media without drawing suspicion [1], [2], [16]. Designing steganographic algorithms for various cover sources is challenging due to the fundamental lack of accurate models. Currently, the most successful approach to design content adaptive steganography is based on minimizing the distortion between the cover and the corresponding stego object. The distortion is obtained by assigning a cost to each modified cover element (e.g., pixel in spatial domain image), and the messages are embedded while minimizing the total distortion which is the sum of costs of all modified elements.

The first method based on the framework of minimizing distortion is HUGO (highly undetectable stego) [3]. HUGO defines the pixel's distortion by the changing amplitude of steganalyzer's features caused by modifying the current pixel, and pixels that make the feature vectors deviated more will have higher costs. The features of steganalyzer SPAM (subtractive pixel adjacency matrix) [4] is used in HUGO. Steganalyzer's features are usually generated by exploiting correlations between the predicted residuals of neighboring pixels [4], [23]. Because the pixels in smooth areas can be accurately predicted, the modifications in such areas will be easily detected by steganalyzers. Therefore the embedding changes of HUGO will be gathered within textured regions. However, HUGO can be detected by steganalyzer with higher dimension of features, such as SRM (spatial rich models) [6].

In SRM, the predicted residuals are generated in various directions and manners, so the correlations between pixels can be further exploited. Therefore, one pixel, to be in a smooth area or a textural area, should be subtly defined for steganography. If the pixel can be accurately modeled in any direction, it should be considered as a smooth point and assigned larger cost. With this insight, Holub et al. proposed the algorithm WOW [5] which assigns high costs to pixels that are more predictable by a bank of directional filters. WOW improves the security of HUGO under the detection of SRM (spatial rich models) [6]. UNIWARD (universal wavelet relative distortion) [7] generalizes the cost function of WOW to make it simpler and more suitable for embedding in an arbitrary domain, including spatial domain and DCT domain for JPEG images. Hence UNIWARD has a similar performance compared to WOW in spatial domain. Li et al. proposed the method HILL [8], which improves WOW by spreading the costs with a low-pass filter. In HILL [8], the local modification probabilities are evened out and thus the modifications cluster in the complex areas.

The above methods design cost function in an ad hoc or empirical manner. Sedighi et al. proposed model-driven approaches [9], [11], in which Multivariate Gaussian (MG) or Multivariate Generalized Gaussian (MVG)'s distribution was used to model noise residuals of pixels by assuming them to be independent but have varying variances. The models are established by estimating the variances and then the costs are computed by minimizing the power of an optimal statistical

test. In fact, small costs will be assigned to residuals modeled with large variances, which just are the highly textured regions.

As summarized in [10], the above adaptive steganographic schemes obey the following two rules:

1) ***Complexity Prior*.** This rule means that the staganographer should give priority for modification to the complex areas that are hard to be modeled. In fact, all the methods in [3], [5], [7]–[9], [11] define cost functions by investigating how to reasonably define the complex degrees of pixels in the sense of resisting detection.
2) ***Cost Spreading*.** Spreading rule means that the costs of modifying two neighboring elements should be similar. In other words, an element with high modification-priority should spread its high-rank to neighborhood, and vice versa. This rule was first successfully used in HILL [8] to improve WOW [5], and then the same idea is used in [12] to improve MVG.

All of the above methods are based on the concept of minimizing the sum of costs of all changed pixels, which is called additive distortion model. In additive distortion model, the modifications on pixels are assumed to be independent. Intuitively, the neighboring embedding changes will interact, thus non-additive distortion model may be more suitable for adaptive steganography. However, it is not clear of how to define costs for non-additive model. Although non-additive elements have been considered in the designs of HUGO, it does not provide stronger undetectability than algorithms based on additive model. Recently, the first effective principle on how to exploit the power of non-additive distortion was found independently by Denemark et al. [14] and Li et al. [13], which implies that synchronizing the modification directions of neighboring pixels can significantly improve the security under detection. The idea used in [14] and [13] can be summarized as the following rule.

3) ***Modification Direction Synchronizing (MDS)*.** This rule is for non-additive distortion model, which means that changing the neighboring pixels in the same directions, i.e., +1 or -1 at the same time, will introduce smaller costs. This rule is also called Clustering Modification Directions (CMD) in [13].

In this paper, we propose a novel rule for improving the security of adaptive steganography from a very different perspective. We notice that many steganographic methods in the framework of minimizing distortion have been presented, and some of them have comparable security performances in resisting detection while they assigning the same pixel with very different modification probabilities because of defining distortion in different manners. We call such pixels as controversial pixels. We consider that these controversial pixels have potential to accommodate more payloads, and thus the undetectability can be improved by giving priority of modifications to such controversial pixels. We call this novel rule as:

4) ***Controversial Pixels Prior (CPP)*.** This rule is for improving several comparative adaptive steganographic methods. According to the CPP rule, the controversial pixels, i.e., the pixels are signed very different modification probabilities by several comparable adaptive steganographic methods, can be given priority for modifications.

With CPP rule, we first improve WOW and UNIWARD who have almost the same performances as shown in Table I. Although HILL [8] is somewhat better than ternary MVG [11] under SRM as shown in Fig. 11, we found that they can also be improved by our proposed CPP rule, especially for large relative payloads, which verifies that our CPP rule is effective.

The rest of this paper is organized as follows. After preliminaries in section II, we make a description on our proposed CPP rule in section III. In section IV and section V, further explorations on the CPP rule have been given and {WOW, UNIWARD}, {HILL, MVG} are taken as two groups of examples to demonstrate the advantages of our CPP rule. The paper is concluded in section VI.

II. The framework of minimal-distortion steganography

In this paper, matrices, vectors and sets are written in boldface, and k-ary entropy function is denoted by $H_k(p_1,\dots,p_k)$ for $\sum_{i=1}^{k} p_i = 1$.

The cover sequence is denoted by $\mathbf{x} = (x_1, x_2, ..., x_n)$, where the signal x_i is an integer, such as the gray value of a pixel. The embedding operation on x_i is formulated by the range I_i. An embedding operation is called binary if $|I_i| = 2$ and ternary if $|I_i| = 3$ for all i. For example, the ± 1 embedding operation is ternary embedding with $I_i = \{x_i - 1, x_i, x_i + 1\}$.

In the model established in [19], the cover $\mathbf{x}$ is assumed to be fixed, so the distortion introduced by changing $\mathbf{x}$ to $\mathbf{y} = (y_1, y_2, ..., y_n)$ can be simply denoted by $D(\mathbf{x}, \mathbf{y}) = D(\mathbf{y})$. Assume that the embedding algorithm changes $\mathbf{x}$ to $\mathbf{y} \in \mathcal{Y}$ with probability $\pi(\mathbf{y}) = P(Y = \mathbf{y})$ which is called modification probability (MP), and thus the sender can send up to $H(\pi)$ bits of message on average with average distortion $E_\pi(D)$ such that

$$H(\pi) = -\sum_{\mathbf{y}\in\mathcal{Y}} \pi(\mathbf{y}) \log \pi(\mathbf{y}),\ E_\pi(D) = \sum_{\mathbf{y}\in\mathcal{Y}} \pi(\mathbf{y}) D(\mathbf{y}). \quad (1)$$

For a given message length L, the sender wants to minimize the average distortion, which can be formulated as the following optimization problems:

$$\min_\pi E_\pi(D), \quad (2)$$

$$\text{subject to } H(\pi) = L. \quad (3)$$

Following the maximum entropy principle, the optimal π has a Gibbs distribution [19]:

$$\pi_\lambda(\mathbf{y}) = \frac{1}{Z(\lambda)} \exp(-\lambda D(\mathbf{y})), \quad (4)$$

where $Z(\lambda)$ is the normalizing factor such that

$$Z(\lambda) = \sum_{\mathbf{y}\in\mathcal{Y}} \exp(-\lambda D(\mathbf{y})). \quad (5)$$

The scalar parameter $\lambda > 0$ can be determined by the payload constraint (3). In fact, as proven in [18], the entropy in (3) is monotone decreasing in λ, thus for a given L in feasible region, λ can be fast determined by binary search.

Specially, if the embedding operations on x_i's are independent mutually, the distortion introduced by changing $\mathbf{x}$ to $\mathbf{y}$ can be thought to be additive, and be measured by $D(\mathbf{y}) = \sum_{i=1}^{n} \rho^{(i)}(y_i)$, where $\rho^{(i)}(y_i) \in \mathbb{R}$ is the cost of changing the ith cover element x_i to y_i ($y_i \in I_i, i = 1, 2 \dots n$). In this case, the optimal π is given by

$$\pi(y_i) = \frac{\exp(-\lambda\rho^{(i)}(y_i))}{\sum_{y_i \in I_i} \exp(-\lambda\rho^{(i)}(y_i))}, \quad i = 1, 2, \dots, n. \tag{6}$$

For additive distortion, there exist practical coding methods to embed messages, such as STCs (Syndrome-Trellis Codes) [19], which can approach the lower bound of average distortion (2).

III. Steganography Based on Rule of CPP

In this section, we will propose a steganographic enhancing method by using the rule of CPP, which generates a new distortion function from several existing ones. As shown in Eq.(6), the distortion can be converted into MP which then determines the payloads assigned to each pixel. Therefore, we will fix our attention on the modification probabilities when searching for controversial-pixels.

The framework of the proposed method is depicted in Fig. 1. Firstly, for a given relative payload, γ bit per pixel (bpp), we set a referenced relative payload γ'. And then we compute MPs with the M existing distortion functions respectively by using Eq.(6) and γ'. Secondly, we label the controversial-pixels according to the MPs. Thirdly, we adjust the MPs with the rule of CPP. Fourthly, the adjusted MPs are converted into a new distortion function. Finally, the messages are embedded with STC according to the new distortion function. Next, we will take ± 1 embedding in spatial images as an example to describe the above processes in details.

The ± 1 embedding is ternary with range $I = \{-1, 0, +1\}$, where 0 means the pixel values keep invariant. If the MP of pixel x_i is p_i, and thus the probability to keep x_i invariant is $\pi_i(0) = 1 - p_i$. We assume that the modifications, $+1$ and -1, have the same probability, i.e., $\pi_i(+1) = \pi_i(-1) = \frac{1}{2}p_i$. In fact, in most adaptive steganogaphic schemes [3], [5], [7]–[9], [11], $+1$ and -1 on a pixel are assigned the same cost so they have the same probability.

Assume that there are M steganographic schemes with comparable performances, and each of them is defined by an additive distortion function D_k for $1 \le k \le M$. The cover is a spatial image consisting of N pixels $\{x_1, \dots, x_N\}$. For the referenced message length $L' = \gamma' N$ and the distortion function D_k, we can calculate the MPs of the N pixels, denoted by $\mathbf{p}_k = \{p_{k,1}, p_{k,2}, ...p_{k,N}\}$, $1 \le k \le M$ with Eq. (6). Collect all the probabilities obtained from the M distortion functions, we get a $N \times M$ matrix $\mathbf{R}$.

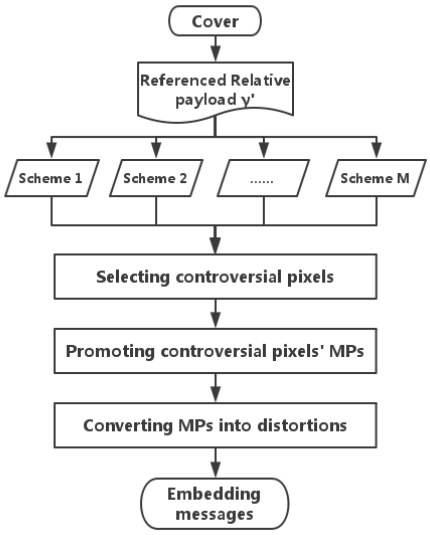


Fig. 1. The flowchart of proposed CPP based method.

$$\mathbf{R} = [\mathbf{r_1}, \mathbf{r_2}, ..., \mathbf{r_N}] = \begin{bmatrix} p_{1,1}, & p_{1,2}, & \dots, & p_{1,N} \\ p_{2,1}, & p_{2,2}, & \dots, & p_{2,N} \\ \vdots & \dots & \vdots & \\ p_{M,1}, & p_{M,2}, & \dots, & p_{M,N} \end{bmatrix}_{N \times M}$$

Here the ith column $\mathbf{r}_i = \{p_{1,i}, p_{2,i}, p_{3,i}, ..., p_{M,i}\}^T$ consists of the M MPs on pixel x_i obtained by the M distortion functions.

In order to describe the divergence among the elements of $\mathbf{r}_i$, we calculate the variance:

$$v_i = \frac{1}{M} \sum_{k=1}^{M} (p_{k,i} - \bar{p}_i)^2, \tag{7}$$

where $\bar{p}_i = \frac{1}{M} \sum_{k=1}^{M} p_{k,i}$ is the mean of all the elements of the ith column. We call v_i the "probability variance" (PV) of the pixel x_i. The bigger the v_i is, the more controversial the modification probability of pixel x_i will be. The pixels with large PV values are those so-called controversial pixels in our scheme. Although these steganographic schemes have similar performances in resisting detection of steganalysis, they assign very different MPs on the controversial pixles, which implies that steganalysis is not sensitive to such pixels. So we can assign more payloads to the controversial pixels by increasing MPs on them.

Denote the $(1 - \alpha)\%$ quantile of $\mathbf{V} = \{v_1, v_2, ...v_N\}$ by T_α. We define the pixels with PVs larger than T_α as the controversial pixels. In other words, the pixels with top $\alpha\%$ PVs are selected as controversial pixels. We call α as controversial threshold and we will discuss how to set α in next section. We fix our attention only on these controversial

pixels and adjust their modification probabilities according to the PVs. To do that, we set

$$v_i' = \begin{cases} v_i & \text{if } v_i > T_\alpha \\ 0 & \text{otherwise} \end{cases} \quad 1 \le i \le N, \tag{8}$$

and adjust the modification probabilities by

$$p_i' = \bar{p}_i e^{v_i'}, 1 \le i \le N. \tag{9}$$

By Eq. (9), we first set the modification probability of x_i to be the mean of $\mathbf{r}_i$. And then, if x_i is a controversial pixels, its MP is increased by multiplying e^{v_i}; otherwise, its MP keeps to be $\bar{p}_i$ because $v_i' = 0$.

Note that, for ± 1 embedding, when the MP $p_i = \frac{2}{3}$, the pixel x_i has the largest average payload $\log_2 3$. Therefore we limit the adjusted MPs by

$$p_i'' = \min\left\{p_i', \frac{2}{3}\right\}, 1 \le i \le N. \tag{10}$$

To embed message with STCs, we should convert the MP to a distortion function. Denote $\pi_i(+) = \pi_i(-) = p_i''/2$ and $\pi_i(0) = 1 - p_i''$. By Eq. (6), the corresponding distortion function $\rho_i(l)$ ($l \in I$) satisfies

$$\pi_i(l) = \frac{\exp(-\lambda\rho_i(l))}{\sum_{t\in I}\exp(-\lambda\rho_i(t))}, l \in I; 1 \le i \le N. \tag{11}$$

To solve $\rho_i(l)$ from Eq. (11), without loss of generality, we can set $\lambda = 1$ because λ is monotone decreasing w.r.t. the message length as proven in [18]. We set distortion as

$$\rho_i(l) = \ln\frac{\pi_i(0)}{\pi_i(l)},\ l \in I, 1 \le i \le N. \tag{12}$$

We call $\rho_i(l)$ in Eq. (12) as adjusted distortion fuction, and it can be easily verified that the adjusted distortion satisfies Eq. (11).

Finally, we obtain a new steganographic algorithm determined by the adjusted distortion function (12), according to which we embed γN bits of message with STCs. Note that the optimal referenced relative payload γ' is usually not equal to the target relative payload γ. We will discuss how to set γ' in the next section. The details of the CPP based method is described in Algorithm 1.

Algorithm 1 CPP Based Steganography

Input: A cover image **x** with N pixels $x_1, \dots x_n$; L bits of message **M** which determines the target relative payload $\gamma = L/N$; M comparable distortion functions for adaptive steganography.

Output: The stego image **y**.

1) Set referenced relative payload γ' and controversial threshold α according to the target relative payload γ.
2) For the M distortion functions, compute the MPs of pixels, $\mathbf{p}_k = \{p_{k,1}, p_{k,2}, ...p_{k,N}\}$, $1 \le k \le M$, by Eq. (6) according to the referenced relative payload γ'.
3) Calculate PV v_i for $1 \le i \le N$ with Eq. (7), and then select the pixels with top $\alpha\%$ PVs as controversial pixels.
4) Adjust the MPs with Eq. (8)-(10), and then calculate the adjusted distortion function with Eq. (12).
5) Embed the L bits of messages **M** into cover image **x** with STCs according to the adjusted distortions, and finally output the stego image **y**.

IV. Improving WOW and UNIWARD with the rule of CPP

In this paper, the security of all steganographic schemes will be evaluated using a steganalyzer that is a detector trained on a given cover source and its stego version embedded with a fixed payload. The detector will be trained by using state-of-the-art 34,671-dimensional SRM feature set [6] with the ensemble classifiers [21]. The performance on resisting detection is evaluated by the testing error which is computed as the mean value of the false positive rate and the false negative rate, averaged over 10 random splits of the data set. Larger classification error rate means stronger security. All the steganalysis experiments are based on the BOSSbase database ver.1.01 [20] containing 10,000 512×512 8-bit gray-scale images coming from eight different cameras.

In this section, we apply the rule of CPP to the steganographic schemes WOW [5] and UNIWARD [7] by setting $M = 2$ in Algorithm 1.

A. Distortion Functions in WOW and UNIWARD

The distortion function of WOW [5] is designed with the help of a group of directional filters, which are denoted by $\boldsymbol{D}^{(k)}(k = 1, \dots n)$. Define a quantity called embedding suitability and denote it by $\xi_{i,j}^{(k)}$. It is computed as the weighted absolute values of the filter residual differences between a cover image and the image after changing only one pixel.

Since the absolute values of the filter residuals are selected as weights, and the filter residual differences have the same form as a rotated directional filter, the embedding suitabilities can be computed by:

$$\boldsymbol{\xi}^{(k)} \triangleq (\xi_{i,j}^{(k)})_{(n_1\times n_2)} = |\boldsymbol{X} \otimes \boldsymbol{D}^{(k)}| \odot |\boldsymbol{D}^{(k)}| \tag{13}$$

The underlying assumption is that if the filter residual is small in one of the directions, the corresponding pixel is predictable, and thus should be assigned a high cost.

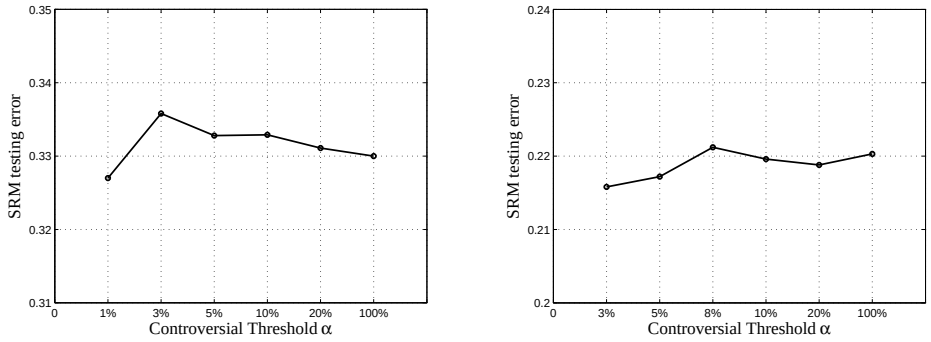
The distortion function of UNIWARD [7] depends on the choice of a directional filter bank and one scalar parameter. The smoothness of a given image $\boldsymbol{X}$ is evaluated along the horizontal, vertical, and diagonal directions by computing the directional residuals $\boldsymbol{W}^{(k)} = \boldsymbol{K}^{(k)} \otimes \boldsymbol{X}$ which has $n_1 \times n_2$ elements.

For a pair of cover and stego images, $\boldsymbol{X}$ and $\boldsymbol{Y}$, their uv_{th} wavelet coefficient in the k_{th} subband of the first decomposition level can be denoted with $\boldsymbol{W}_{uv}^{(k)}(\boldsymbol{X})$ and $\boldsymbol{W}_{uv}^{(k)}(\boldsymbol{Y})$, $k = 1, 2, 3, u \in \{1, \dots, n_1\}, v \in \{1, \dots, n_2\}$. The UNIWARD

(a) Full-size image 1013.pgm (b) $\alpha = 3$

(c) $\alpha = 5$ (d) $\alpha = 10$

Fig. 3. The location of the top $\alpha\%$ controversial pixels to a sample cover image (a) with $\alpha = 3$ (b), $\alpha = 5$ (c),and $\alpha = 10$ (d) respectively, where the dark points represent the controversial pixels and the gray points represent general ones, respectively.



(a) The testing error–α curve for 0.2bpp (b) The testing error–α curve for 0.4bpp

Fig. 4. The change trends of testing error w.r.t. α at relative payload 0.2 bpp and 0.4 bpp when applying CPP to WOW and UNIWARD.

distortion function is the sum of relative changes of all wavelet coefficients with respect to the cover image:

$$D(\boldsymbol{X},\boldsymbol{Y}) \triangleq \sum_{k=1}^{3}\sum_{u=1}^{n_1}\sum_{v=1}^{n_2}\frac{|\boldsymbol{W}_{uv}^{(k)}(\boldsymbol{X}) - \boldsymbol{W}_{uv}^{(k)}(\boldsymbol{Y})|}{\sigma + |\boldsymbol{W}_{uv}^{(k)}(\boldsymbol{X})|}, \tag{14}$$

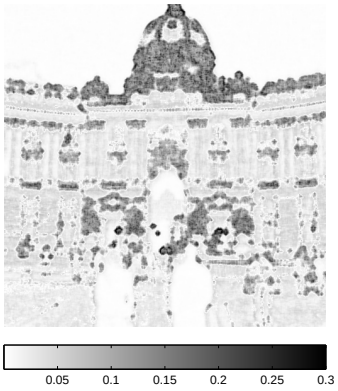
where $\sigma > 0$ is a constant stabilizing the numerical calculations.

Obviously, the distortion function of UNIWARD bears some similarity to WOW in the sense that the embedding costs of those two steganography are both computed from three directional residuals. Therefore, UNIWARD and WOW have comparative performance in resisting the detection of SRM as shown in Table I, which satisfies the assumption of the rule of CPP.

B. Setting Controversial Threshold α

To apply Algorithm 1, we only need to set the parameters α and γ'. To discuss how to esitmate the optimal α, we first

(a) Full-size image 1013.pgm



(b) The values of probability variances for CPP

Fig. 2. The values of probability variances for improving WOW and UNIWARD with CPP (0.4 bpp).

TABLE I
TESTING ERRORS OF SRM FOR DETECTING WOW AND UIWARD.

Relative Payload (bpp)	0.05	0.1	0.2	0.3	0.4	0.5
UNIWARD	0.454	0.402	0.324	0.261	0.207	0.164
WOW	0.457	0.407	0.322	0.263	0.213	0.169

set $\gamma' = \gamma$ in the experiments of this subsection, i.e., set the referenced relative payload equal to the given relative payload.

In Fig. 2, we visualize the values of PV, v_i's, which shows that the most controversial-pixels cluster in complex area. On the other hand, we give Fig. 3 to observe the distribution of the controversial pixels for setting different threshold α, which shows that controversial pixels may spread to smooth areas when α is large. When embedding a short message, the modifications in complex regions are enough to carry the payload. Therefore, we guess that, for a small relative payload, we only need small threshold α, which is verified by following experimental results. Fig. 4(a) shows that the largest testing error appears at $\alpha = 2$ for relative payload $\gamma = 0.2$ while Fig. 4(b) shows that the largest testing error appears at $\alpha = 8$ for $\gamma = 0.4$.

We list the optimal α for different relative payload γ in Table II. To observe the relation between α and γ, we make a scatter plot in Fig. 5 for the values of Table II , which shows the optimal α increases with γ and they have a strongly linear relation. Therefore we establish the following linear regression model between α and γ.

TABLE II
THE OPTIMAL PARAMETER α FOR IMPROVING WOW AND UNIWARD.

Relative Payload (bpp)	0.05	0.1	0.2	0.3	0.4	0.5
Optimal α	0.8	1.7	3.8	6.1	8.5	11.4

$$\alpha = 23.38\gamma - 0.67, \tag{15}$$

by which we can estimate the optimal α for a given relative payload γ.

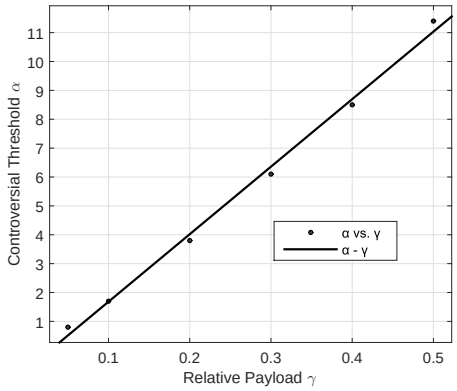


Fig. 5. The scatter plot of (γ, α) for improving WOW and UNIWARD with CPP.

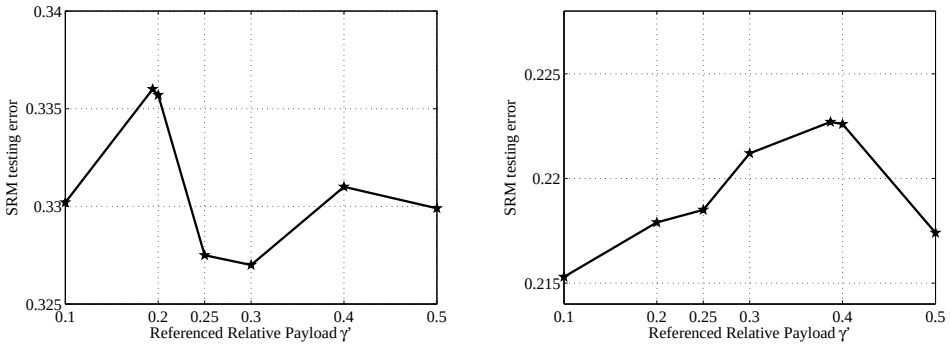
C. Setting Referenced Relative Payload γ'

In Algorithm 1, from a referenced relative payload γ', we get an adjusted distortion function Eq. (12), by which we can embed message with any relative payload belonging to a reasonable range. Therefore, the question is what is the optimal γ' for a given relative payload γ? In the above subsection, we set $\gamma' = \gamma$. Next, we will discuss whether it is just the optimal setting.

Fig. 6(a) and Fig. 6(b) shows how the SRM's testing errors changes with γ' for a given relative payload γ. Table III lists the values of optimal γ' for each given γ, which shows that the optimal γ' is usually a little smaller than γ. And Fig. 6 also show that, by setting $\gamma' = \gamma$, we can get a sub-optimal solution which is very close to the optimal solution. Therefore, for simplicity, we proposed to set $\gamma' = \gamma$.

TABLE III
THE OPTIMAL γ' FOR DIFFERENT γ (*bpp*) FOR IMPROVING WOW AND UNIWARD.

Relative Payload γ	0.1	0.2	0.3	0.4	0.5
Optimal γ'	0.097	0.194	0.290	0.387	0.484



(a) 0.2 bpp (b) 0.4 bpp

Fig. 6. The change trends of testing errors w.r.t. referenced relative payload γ' for different relative payloads.

D. Performance Comparison

In this section, we compare the security of WOW, spatial UNIWARD and the method improved by the rule of CPP. When applying CPP, we set $\gamma' = \gamma$ and determine α by Eq. (15) for a given relative payload γ. The results are depicted in Fig. 7, which shows that CPP can enhance the ability of resisting detection of WOW and UNIWARD for various relative payloads.

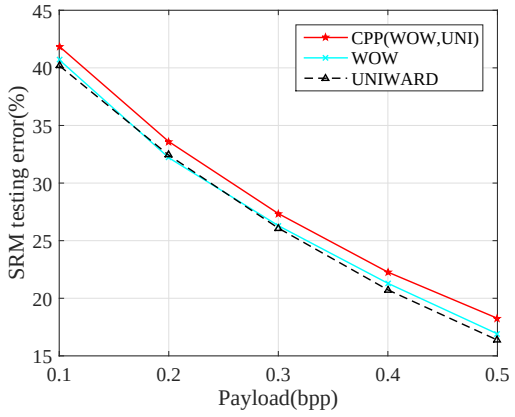


Fig. 7. Comparison between WOW, UNIWARD and the improved method by CPP under detection of SRM.

V. Improving HILL and MVG with the Rule of CPP

In this section, we apply CPP to HILL [8] and MVG [11]. Note that Sedighi et al. recently proposed an improved version of MVG in [11]. In this paper,we use the old version of MVG [11], because of the following reason. The performance of HILL [8] and MVG in [11] are not so close like that of WOW and UNIWARD, which departures from the assumption of CPP such that the M candidates of steganographic algorithms are comparable. By this example, we want to show that the principle of CPP still can enhance the security even if the performance of the existing steganographic algorithms are not so consistent.

A. Cost Function in HILL and MVG

HILL [8] is an improved method based on WOW. The definition of distortions in HILL follows the *spreading rule* (*or clustering rule*) [10]. The cost value of a pixel is weighted by its neighboring cost values, and the mutual dependencies among cost values are taken into consideration.

In HILL, the directional filter $\boldsymbol{D}^{(k)}$ can be replaced by any high-pass filter $\boldsymbol{H}^{(k)}$, regardless of whether directional or non-directional, to locate the less predictable area. And since the elements in the second term $|\boldsymbol{D}^{(k)}|$ are all non-negative, the filter $|\boldsymbol{D}^{(k)}|$ can be substituted with a low-pass filter to make it more flexible for use. Besides, when the low pass-filter is central symmetric, correlation can be replaced by convolution.

Thus, the new designed embedding suitability is no longer the weighted filter residual difference, but the smoothed filter residual. It can be interpreted as using a high-pass filter and then a low-pass filter to locate the less predictable regions. Hence the process of obtaining cost value can be further simplified as:

$$\boldsymbol{\varrho} = \frac{1}{|\boldsymbol{X} \otimes \boldsymbol{H}^{(1)}| \otimes \boldsymbol{L}_1} \otimes \boldsymbol{L}_2 \tag{16}$$

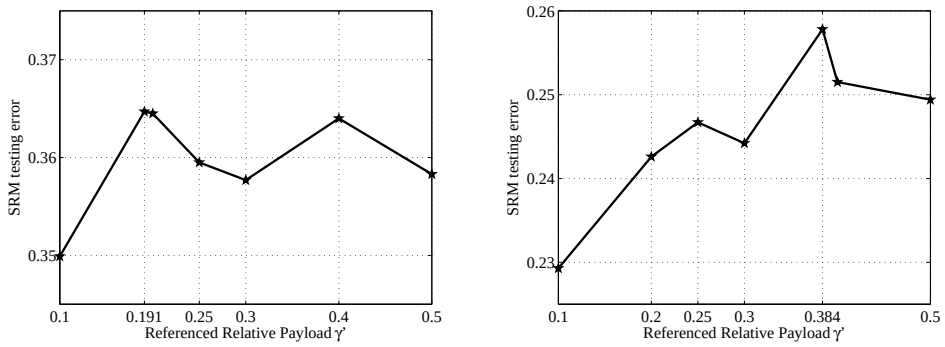
where $\boldsymbol{L}_1$ and $\boldsymbol{L}_2$ are two low-pass filters.

MVG [11] is entirely model driven. In the approach of MVG, modification probabilities of cover pixels are derived from the cover model to minimize the power of an optimal statistical test. MVG can be summarized as two steps: *First*, the sender estimates the cover model parameters, the pixel variances, when modeling pixels as a sequence of independent but not identically distributed generalized Gaussian random variables. *Second*, the modification probabilities for changing each pixel are computed by solving a pair of non-linear algebraic equations for minimizing the total KL divergence between the cover object and stego object.

B. Setting Parameters α and γ'

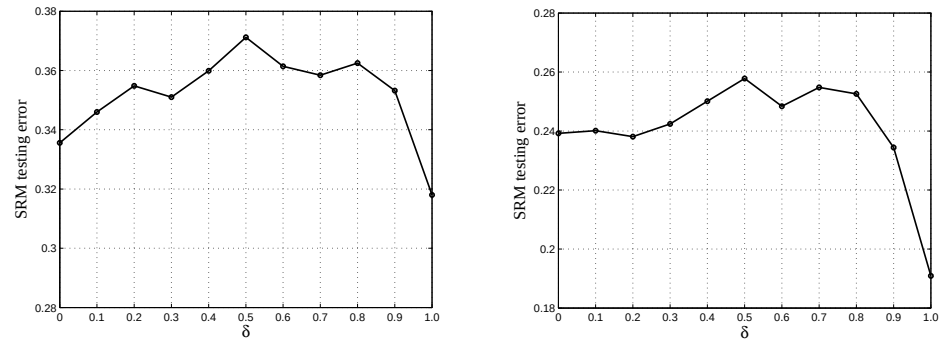
To apply CPP to HILL and MVG, we also need to determine the parameters α and γ' in Algorithms 1.

To determine α, we give the scatter plot of optimal α for corresponding relative payload γ in Fig. 8, which also shows



(a) 0.2 bpp (b) 0.4 bpp

Fig. 9. The change trends of testing errors with respect to γ' at relative payload 0.2 bpp and 0.4 bpp respectively.



(a) 0.2 bpp (b) 0.4 bpp

Fig. 10. The change trends of testing errors with respect to weights δ at relative payload 0.2 bpp and 0.4 bpp respectively.

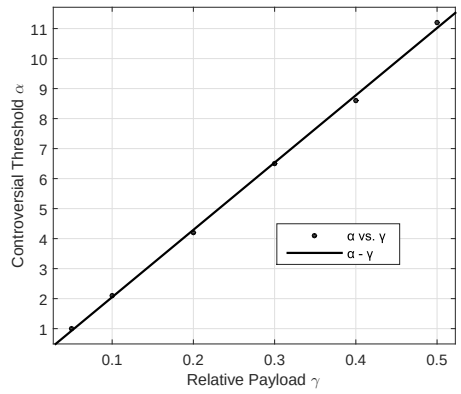


Fig. 8. The scatter plot of (γ, α) for improving HILL and MVG with CPP.

a strong linear relation. By the values depicted in Fig. 8, we establish the following linear regression model to estimate α.

$$\alpha = 22.42\gamma - 0.19. \tag{17}$$

It is interesting that, the model (17) is so similar to the model (15). Both of these two linear models have intercepts very close to zero and slopes close to twenty.

To determine γ', we list the optimal γ' for various relative payload γ under the detection of SRM in Table IV, which shows that the optimal γ' increases with γ. And Fig.9 shows that, when $\gamma' = \gamma$, we can get a nearly optimal performance, so we also set $\gamma' = \gamma$ as what we have done for improving WOW and UNIWARD.

C. Discussion on Basic Modification Probability

In Eq. (9), we first set the basic modification probability as $\bar{p}_i = \frac{1}{M}\sum_{k=1}^{M} p_{k,i}$, which is the average modification probability obtained by several comparable steganographic algorithms, and then we adjust the modification probability according to PV. The assumption behind such setting is that the performance of the M steganographic algorithms are comparable. But in the present example, HILL outperforms MVG as shown in Fig. 11. Therefore, intuitively, we should compute the basic modification probability with weights biased to HILL. In other words, if the modification probabilities obtained by HILL and MVG are denoted by p_i^{HILL} and p_i^{MVG} respectively, we should calculate $\bar{p}_i$ by

$$\bar{p}_i = \delta p_i^{HILL} + (1-\delta) p_i^{MVG}, \tag{18}$$

with $\delta > 0.5$.

We depict the changing trends of testing errors along with the weight δ for relative payload 0.2 bpp and 0.4 bpp respectively in Fig. 10, which shows that the value of δ will

significantly influence the performance, but the optimal setting is also $\delta = 0.5$ although HILL and MVG in [11] are not so comparable. Therefore, we still use Eq. (9) to calculate modification probability when applying CPP to HILL and MVG.

TABLE IV
THE OPTIMAL γ' OF DIFFERENT γ (*bpp*) FOR IMPROVING HILL AND MVG.

Relative Payload γ	0.1	0.2	0.3	0.4	0.5
Optimal γ'	0.096	0.191	0.288	0.384	0.482

D. Performance Comparison

We compare the performance of HILL, MVG and the corresponding CPP enhanced method for resisting the detection of SRM in Fig. 11, which shows that the security of HILL and MVG can also be improved by the rule of CPP. Therefore, we conclude that the rule of CPP will be effective as long as the performance of the candidates of steganographic algorithms are close to each other. On the other hand, the promotion of performance is not as significant as it for WOW and UNIWARD, which implies that the assumption of "performance being comparable" is important for CPP. In fact, because there exist some differences between the performances of HILL and MVG, the concept of controversial-pixels defined in this paper maybe not so reasonable in this case.

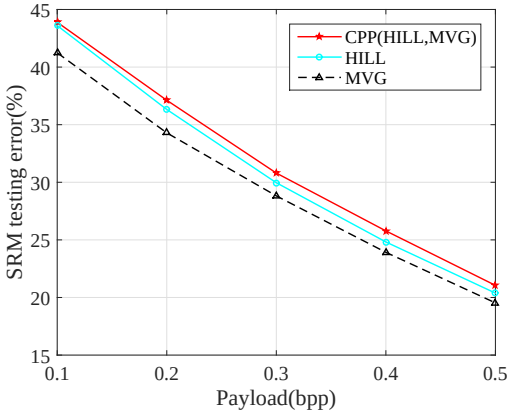


Fig. 11. Comparison between HILL and MVG and the improved method with CPP under the detection of SRM.

VI. CONCLUSION

Currently, the most effective model for adaptive steganography is to embed messages while minimizing a carefully defined distortion function. In recent advances, a novel steganographic method is usually obtained from an existing one by modifying the latter distortion function according to some rules. For instance, HILL [8] can be viewed as an improved version of WOW [5] by smoothing its distortion function with the spreading rule. The non-additive distortion based methods [13], [14] are obtained by modifying the additive distortion function defined in MVG or HILL according to the rule of modification direction synchronizing.

In the present paper, we proposed the rule of Controversial-Pixels Prior, with which we generate a new steganographic distortion function by modifying a group of previous ones that have comparable performance. Taking {WOW, UNIWARD} and {HILL, MVG} as two pairs of examples, we show that the rule of CPP can improve the security of state of the art steganographic algorithms.

So far, we only use the CPP rule to improve steganography in spatial images with additive distortion for ± 1 embedding. In out future work, we will try to apply CPP rule to steganography in other cover formats such as JPEG images or video and other modification manners such as binary or pentary embedding. Furthermore, a more interesting direction is to improve non-additive distortion based steganography with the rule of CPP.

ACKNOWLEDGMENT

This work was supported in part by the Natural Science Foundation of China under Grant 61572452 and Grant 61502007, in part by the China Postdoctoral Science Foundation under Grant 2015M582015, and in part by the Strategic Priority Research Program of the Chinese Academy of Sciences under Grant XDA06030601.

REFERENCES

[1] J. Fridrich, Steganography in Digital Media: Principles, Algorithms and Applications. Cambridge University Press, 2009.
[2] B. Li, J. He, J.w. Huang, and Y. Q. Shi , "A survey on image steganography and steganalysis," Journal of Information Hiding and Multimedia Signal Processing, vol. 2, no. 2, pp. 142-172, 2011.
[3] T. Pevny, T.Filler, and T. Bas, "Using high-dimensional image models to perform highly undetectable steganography," Proc. of International Workshop on Information Hiding, vol. LNCS 6387, pp. 161-177, Jun. 28-30, 2010.
[4] T. Pevny, P. Bas, and J. Fridrich, "Steganalysis by subtractive pixel adjacency matrix," IEEE Trans. on Inf. Forensics and Security , vol. 5, no. 2, pp. 215-224, Jun. 2010.
[5] V. Holub and J. Fridrich, "Designing steganographic distortion using directional filters," Proc. of IEEE Workshop on Information Forensic and Security, pp. 234-239, Dec. 2-5, 2012.
[6] J. Fridrich and J. Kodovsky, "Rich models for steganalysis of digital images ," IEEE Trans. on Inf. Forensics Security, vol. 7, pp. 868-882, Jun. 2012.
[7] V. Houlb and J. Fridrich, "Digital image steganography using universal distortion," Proc. of ACM Workshop on Information hiding and multimedia security, pp.59-68, Jun. 17-19, 2013.
[8] B. Li, M. Wang, J. W. Huang, and X. L. Li, "A new cost funtion for spatial image steganpgraphy," Proc. of IEEE International Conference on Image Processing, Oct. 27-30, 2014.
[9] J. Fridrich and J. Kodovsky, "Multivariate Gaussian model for designing additive distortion for steganography," Proc. of IEEE ICASSP, Vancouver, BC, May 26-31, 2013.
[10] B. Li, S. Tan, M. Wang, and J.W. Huang, "Investigation on cost assignment in spatial image steganography," IEEE Trans. on Inf. Forensics Security, vol. 9(8), pp. 1264-1277, May, 2014.
[11] V. Sedighi, J. Fridrich, and R. Cogranne, "Content-adaptive pentary steganography using the multivariate generalized gaussian cover model," Proc. of SPIE, Electronic Imaging, Media Watermarking, Security, and Forensics 2015, vol. 9409, San Francisco, CA, Feb. 8-12, 2015.
[12] V. Sedighi, R. Cogranne, and J. Fridrich, "Content-adaptive steganography by minimizing statistical detectability," IEEE Trans. on Inf. Forensics Security, vol. 11(2), pp. 221-234, Oct. 2015.

[13] B. Li, M. Wang, X.L. Li,and S. Tan, "A strategy of clustering modification directions in spatial image steganography," IEEE Trans. on Inf. Forensics Security, vol. 10, pp.1905 - 1917, May, 2015.
[14] T. Denemark and J. Fridrich, "Improving steganographic security by synchronizing the selection channel," Proc. of 3rd Workshop on I-H&MMSec, Portland, Oregon, Jun. 17-19, 2015.
[15] T. Filler and J. Fridrich, "Design of adaptive steganographic schemes for sigital images," Proc. of SPIE, Electronic Imaging, Media Watermarking, Security, and Forensics XIII, San Francisco, CA, Jan. 24-26, 2011.
[16] T. Pevny and J. Fridrich, "Benchmarking for steganography," Proc. of 10th International Workshop on Information Hiding, vol. LNCS 5284, pp. 251-267, Springer Berlin Heidelberg, May 19-21, 2008.
[17] J. Fridrich and T. Filler, "Practical methods for minimizing embedding impact in steganography," Proc. of SPIE, Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents IX, vol. 6505, p. 650502, 2007.
[18] T. Filler and J. Fridrich, "Gibbs construction in steganography," IEEE Trans. on Inf. Forensics Security, vol. 5, no. 4, pp. 705C720, Sept. 2010.
[19] T. Filler, J. Judas, and J.Fridrich, "Minimizing additive distortion in steganography using syndrome trellis codes," IEEE Trans. on Inf. Forensics Security, vol. 6, no. 3-2, pp. 920-935, Apr. 2010.
[20] P. Bas, T. Filler, and T. Pevny, "Break our steganographic system - the ins and outs of organizing boss," Proc. of 13th International Workshop on Information Hiding, vol. LNCS 6958, pp. 59-70, Springer Berlin Heidelberg, May 18-20, 2011.
[21] J. Kodovsky, J. Fridrich, and V. Holub, "Ensemble classifiers for steganalysis of digital media," IEEE Trans. on Inf. Forensics Security, vol. 7, no. 2, pp. 432-444, Nov. 2012.
[22] T. Denemark, V. Sedighi, V. Holub, R. Cogranne, and J. Fridrich, "Selection-channel-aware rich model for steganalysis of digital images," Proc. of 6th IEEE International Workshop on Information Forensics and Security, Atlanta, GA, USA, Dec. 3-5, 2014.
[23] J. Kodovsky, T. Pevny, and J. Fridrich, "Modern Steganalysis Can Detect YASS," Proc. of SPIE, Electronic Imaging, Media Forensics and Security XII, pages 2 1-2 11, San Jose, CA, Jan. 17-21, 2010.