This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2587388, IEEE Transactions on Circuits and Systems for Video Technology

Decomposing Joint Distortion for Adaptive Steganography

Weiming Zhang, Zhuo Zhang, Lili Zhang, Hanyi Li, Nenghai Yu

***Abstract*—Recent advances on adaptive steganography imply that the security of steganography can be improved by exploiting the mutual impact of modifications between adjacent cover elements such as pixels of images, which is called non-additive distortion model. In this paper, we propose a framework for non-additive distortion steganography by defining joint distortion on pixel blocks. To reduce the complexity for minimizing joint distortion, we design an coding method to decompose the joint distortion (abbreviated to DeJoin) into distortion on individual pixels and thus the message can be efficiently embedded with syndrome trellis codes (STCs). We prove that DeJoin can approach the lower bound of joint distortion. As an example, we define joint distortion according to the principle of Synchronizing Modification Direction (SMD) and then design steganographic algorithms with DeJoin. The experimental results show that the proposed method outperforms previous non-additive distortion steganography when resisting state-of-the-art steganalysis.**

***Index Terms*—Covert communication, steganography, non-additive distortion, joint distortion**

I. Introduction

STEGANOGRAPHY is a technique for covert communication, which aims to hide secret messages into ordinary digital media without drawing suspicion [1]–[4]. Currently, most approaches on content adaptive steganography are based on the model of minimizing distortion between the cover and the corresponding stego object.

Most adaptive steganographic methods adopted additive distortion fuchtions, such as HUGO [5], WOW [6], UNIWARD [7], HILL [8], MVG [9], and MiPOD [10] in which the distortion is defined by assigning costs to individual cover elements. In additive distortion model, the modifications on pixels are assumed to be independent and thus minimizing the overall costs is equivalent to minimizing the sum of costs of individual changed elements. For additive distortion based methods, the practical message embedding is usually realized by the efficient coding method, syndrome-trellis codes (STCs) [11], which can approach the lower bound of average embedding distortion for additive model.

Intuitively, the changes on adjacent pixels will interact, and thus non-additive distortion model will be more suitable for adaptive steganography. Recently, Li et al. [12] and Denemark et al. [13] independently found an effective principle for exploiting the mutual impact of adjacent modifications, which implies that synchronizing modification directions (SMD) of adjacent pixels can significantly improve the performance for resisting detection. In [13], ± 1 modifications are used for embedding messages in spatial images, and firstly each pixel is assigned an initial ± 1 cost with an additive distortion function such as the one used in MVG [9] or HILL [8], and then adjacent pixels are divided into two non-overlapped sub-sets and correspondingly the embedding procedure consists of two rounds. In the second round, the initial distortion on each pixel of the second sub-set will be updated according to the change directions of its neighboring pixels that have been changed in the first round. Distortion updating is based on the principle of SMD and the costs on changes in the same direction will be decreased. Therefore, we call the strategy used in [12], [13] as "updating distortion" (abbreviated to UpDist).

In general, it is still difficult to minimize non-additive distortion for embedding message with low computational complexity. Ker et al. [14] made a comprehensive review of current art in both steganography and steganalysis and then listed a series of open problems, in which the first open problem for steganography is just how to design efficient coding schemes for non-additive distortion functions. In the present paper, we try to solve this problem via a decomposition coding scheme.

We propose a novel framework to exploit interactions between modifications of adjacent pixels by defining joint distortion on pixel blocks. To minimize the joint distortion in practical embedding with low complexity, we propose a method to decompose joint distortion, so this method is called DeJoin for short. By DeJoin, the joint distortion is decomposed into additive distortion on individual pixels and thus the message can be embedded with STC efficiently. We prove that DeJoin can approach the lower bound of average joint distortion for a given payload. As an example, we define joint distortion by using the principle of SMD, and then apply DeJion to embed messages. The experimental results show that DeJoin can improve the performance of the methods in [12], [13].

The rest of this paper is organized as follows. In Section II, we briefly introduce the model of minimal distortion steganography. The framework of decomposing joint distortion is elaborated in Section III. Some examples on how to define joint distortion and the experiments for resisting steganalysis

This work was supported in part by the Natural Science Foundation of China under Grant 61170234 and Grant 61572452, in part by the Strategic and Piloted Project of CAS under Grant XDA06030601, and in part by the project funded by the Priority Academic Program Development of Jiangsu Higer Education Institutions.

All the authors are with CAS Key Laboratory of Electromagnetic S-pace Information, University of Science and Technology of China, Hefei, 230026, China; (e-mail: zhangwm@ustc.edu.cn, zhang589@mail.ustc.edu.cn, ll0414@mail.ustc.edu.cn, lihy1015@mail.ustc.edu.cn, ynh@ustc.edu.cn).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2587388, IEEE Transactions on Circuits and Systems for Video Technology

are given in Section IV. The paper is concluded with a discussion in Section V.

II. Model on Minimal Distortion Steganography

In this paper, matrices, vectors and sets are written in boldface, and k-ary entropy function is denoted by $H_k(p_1,\ldots,p_k)$ 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. An embedding operation is called binary if $|I| = 2$ and ternary if $|I| = 3$ for all i. For example, the ±1 embedding operation is ternary embedding with $I = \{0, +1, -1\}$, where 0 denotes no modification.

In the model established in [11], 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 $\pi(\mathbf{y}) = P(Y = \mathbf{y})$, and thus the sender can send up to $H(\pi)$ bits of message on average with average distortion $E_\pi(D)$ such that

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

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

$$\min\nolimits_\pi E_\pi(D), \tag{2}$$

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

Following the maximum entropy principle, the optimal π has a Gibbs distribution [11]:

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

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

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

The scalar parameter $\lambda > 0$ can be determined by the payload constraint (3). In fact, as proven in [15], the entropy in (3) is monotone decreasing in λ, so 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 ($y_i \in I + x_i, i = 1, 2 \ldots n$). In this case, the optimal π is given by

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

For additive distortion, there exist practical coding methods to embed messages, such as STCs (Syndrome-Trellis Codes) [11], which can approach the lower bound of average distortion (2). STCs are a kind of syndrome coding methods, which can be formulated as

$$Emb(\mathbf{x},\mathbf{m}) = \arg\min_{\mathbf{y}\in C(\mathbf{m})} D(\mathbf{x},\mathbf{y}),\ \text{ and } \mathbf{Hy} = \mathbf{m}. \tag{7}$$

Here, $\mathbf{m}$ is the message vector. $\mathbf{H}$ is the parity-check matrix and $C(\mathbf{m})$ is the coset corresponding to syndrome $\mathbf{m}$. The syndrome coding process is to find the stego vector $\mathbf{y}$, satisfying $\mathbf{Hy} = \mathbf{m}$ and having minimal distortion. After receiving $\mathbf{y}$, the receiver can easily extract messages by computing $\mathbf{Hy}$.

The parity check matrix $\mathbf{H}$ used in STCs is constructed by a $h\times w$ sub-matrix $\hat{\mathbf{H}}$. For STCs, each solution can be represented as a path through the syndrome trellis of $\mathbf{H}$. The height h of the sub-matrix determines the number of paths, and there are k^h choices in each grid of the trellis for k-ary embedding. Therefore, larger h means more powerful capacity to minimize distortion but also higher computational complexity. On the other hand, the complexity of STC will exponentially increase with the number of the modification patterns $k = |I|$. Therefore, although STC can be fast implemented for binary embedding, it cannot be directly extended to multi-ary embedding such as ±1 embedding.

However, ±1 embedding is more suitable for steganography in spatial images. Zhang et al. [16], [17] presented an efficient double-layered embedding scheme, by which ±1 embedding is decomposed into two binary embedding operations, namely embedding messages into LSB (Least Significant Bit) layer and second LSB layer respectively. Motivated by [16], [17], Filler et al. generalized double-layered embedding and thus proposed double-layered STCs for ±1 embedding [11] that is used by most adaptive steganographic methods [6]–[10], [12], [13].

III. Framework of Decomposing Joint Distortion

We take spatial images and ±1 embedding as an example to describe the proposed method, which in general can be replaced with other kinds of covers and embedding manners.

Previous adaptive steganography usually defines distortion function ρ_i on single pixel x_i for $i = 1, \ldots, n$. To consider the interactive impact of modifications between neighboring pixels, we propose to define distortion on pixel blocks. To do that we first divide the cover image into N non-overlapped blocks with size $n_1\times n_2$. Without loss of generality, we assume that $N = n/(n_1 \times n_2)$. For each block B_i ($1 \le i \le N$), we define a joint distortion function $\rho^{(i)}$, according to which we embed messages and minimize the sum of distortion of all blocks. In this model, the distortion of inter-blocks is still additive, but we cannot directly use STCs, because the number of modification patterns in each block is large that causes high computational complexity. To reduce the complexity of embedding processes, we propose a decomposition coding method, which decomposes the joint distortion on blocks into distortions on individual pixels, and thus STCs can be used to embed message efficiently. The proposed method is abbreviated to DeJoin.

We take the 1×2 block as an example to describe the decomposition coding method, i.e., each block consists of two pixels. Assume the block sequence of the cover is

$$B_1 = (x_{1,1}, x_{1,2}), \ldots, B_i = (x_{i,1}, x_{i,2}), \ldots, B_N = (x_{N,1}, x_{N,2}). \tag{8}$$

Because each pixel has three possible modification patterns for ±1 embedding, the 2-pixel block has nine modification

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2587388, IEEE Transactions on Circuits and Systems for Video Technology

patterns. Therefore, the joint distortion function $\rho^{(i)}$ on the ith block includes nine variables, which can be described as a 3×3 matrix with element $\rho^{(i)}(l, r)$ for $(l, r) \in I^2$, where $I = \{+1, -1, 0\}$. Herein, $\rho^{(i)}(l, r)$ denotes the distortion introduced by modifying $(x_{i,1}, x_{i,2})$ to $(x_{i,1} + l, x_{i,2} + r)$ for $(l, r) \in I^2$. We will discuss how to define $\rho^{(i)}(l, r)$ in the next section.

For a given message length L, with Eq. (6) the optimal joint modification probability $\pi^{(i)}$ on the ith block is given by

$$\pi^{(i)}(l, r) = \frac{\exp(-\lambda\rho^{(i)}(l, r))}{\sum_{(u,v)\in I^2} \exp(-\lambda\rho^{(i)}(u, v))},\ (l, r) \in I^2,\ 1 \le i \le N, \tag{9}$$

which satisfies $L = \sum_{i=1}^{N} H_9(\pi^{(i)})$. To reduce complexity of the embedding processing, we design a two-round embedding strategy by decomposing the joint probability $\pi^{(i)}$ into margin probability and conditional probability.

Denoting the margin probability on the first pixel $x_{i,1}$ of the ith block by $\pi_1^{(i)}$, we have

$$\pi_1^{(i)}(l) = \sum_{r\in I} \pi^{(i)}(l, r),\ l \in I. \tag{10}$$

We use $\pi_{2|\,l}^{(i)}(r)$ to denote the probability such that $x_{i,2}$ is changed to $x_{i,2} + r$ under the condition of $x_{i,1}$ having been changed to $x_{i,1} + l$. The conditional probability is calculated by

$$\pi_{2|\,l}^{(i)}(r) = \frac{\pi^{(i)}(l, r)}{\pi_1^{(i)}(l)},\ r \in I, l \in I. \tag{11}$$

In the first round, we will embed

$$L_1 = \sum_{i=1}^{N} H_3\left(\pi_1^{(i)}\right) \tag{12}$$

bits of message into $x_{1,1}, \ldots, x_{i,1}, x_{N,1}$. In the second round, we will embed

$$L_2 = \sum_{i=1}^{N} \sum_{l\in I} \pi_1^{(i)}(l) H_3\left(\pi_{2|\,l}^{(i)}\right) \tag{13}$$

bits of message into $x_{1,2}, \ldots, x_{i,2}, x_{N,2}$. By chain rule, we have

$$L_1 + L_2 = \sum_{i=1}^{N} H_9(\pi^{(i)}) = L. \tag{14}$$

Herein, H_3 is the ternary entropy function and H_9 is the nine-ary entropy function because both $\pi_1^{(i)}$ and $\pi_{2|\,l}^{(i)}$ are ternary probability mass functions (PMF) while $\pi^{(i)}$ is a nine-ary PMF.

For practical embedding, we will use double-layered STCs [11] to embed message in each round. To do that, we transform the modification probability to a corresponding distortion function. In the first round, we transform the margin probability $\pi_1^{(i)}(l)$ to a distortion function on $x_{i,1}$ which is denoted as $\rho_1^{(i)}(l)$ for $l \in I$. By Eq. (6), we have

$$\pi_1^{(i)}(l) = \frac{\exp(-\lambda\rho_1^{(i)}(l))}{\sum_{t\in I} \exp(-\lambda\rho_1^{(i)}(t))}, l \in I; 1 \le i \le N. \tag{15}$$

To solve $\rho_1^{(i)}(l)$ from Eq. (15), without loss of generality, we can set $\lambda = 1$ because λ is monotone decreasing w.r.t. the message length as proven in [15]. We define distortion by

$$\rho_1^{(i)}(l) = \ln \frac{\pi_1^{(i)}(0)}{\pi_1^{(i)}(l)},\ l \in I, 1 \le i \le N. \tag{16}$$

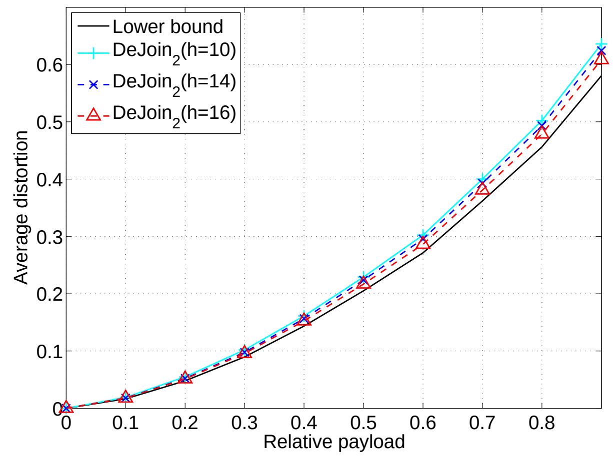


Fig. 1. Performance of DeJoin_2 for minimizing joint distortion.

It can be easily verified that, if we embed message according to the distortion function (16), the optimal modification probability for the message length L_1 is just $\pi_1^{(i)}(l)$ for $l \in I$ and $1 \le i \le N$, because the optimal modification probability is given by setting $\lambda = 1$ in Eq. (15).

After assigning the distortion function (16) to $x_{i,1}$ for $1 \le i \le N$, we apply ± 1 STCs to embed L_1 bits of message into $x_{1,1}, \ldots, x_{N,1}$, by which $x_{i,1}$ will be modified approximatively according to the probability $\pi_1^{(i)}(l)$ for $l \in I$.

In the second round, we embed messages into $x_{1,2}, \ldots, x_{N,2}$ according to the conditional modification probability $\pi_{2|\,l}^{(i)}$ on $x_{i,2}$ for $1 \le i \le N$. Which condition l should be used is determined by the actual embedding result in the first round. For instance, if the $x_{i,1}$ is modified by $+1$ in the first round, we will use the probability mass function $\pi_{2|+1}^{(i)}(r)$ $(r \in I)$ for $x_{i,2}$. After transforming the conditional probability to a distortion function as done in the first round, we embed L_2 bits of message into $x_{1,2}, \ldots, x_{N,2}$ with ± 1 STCs.

At the receiver side, the message can be easily extracted by applying the decoding algorithm of ± 1 STC to the two subsequences respectively.

If we can realize optimal ± 1 embedding in each round, the probability for modifying $(x_{i,1}, x_{i,2})$ to $(x_{i,1} + l, x_{i,2} + r)$ in the above process is equal to

$$\pi_1^{(i)}(l) \times \pi_{2|\,l}^{(i)}(r) = \pi^{(i)}(l, r),\ (l, r) \in I^2, 1 \le i \le N. \tag{17}$$

Therefore, the average distortion introduced by DeJoin reaches the lower bound because $\pi^{(i)}(l, r)$ is the optimal modification probability for embedding L bits under the joint distortion $\rho^{(i)}(l, r)$. Thus, we conclude that, DeJoin can minimize the joint distortion defined on pixel blocks only if STCs can approach the lower bound of average distortion on individual pixels.

We denote the algorithm for 2-pixel blocks by DeJoin_2. To verify the performance of the proposed method, we define joint distortion with random numbers on 2-pixel blocks of a 512×512 image, and then embed messages with DeJoin_2 by setting the parameter h of STCs as $h = 10, 14$ and 16 respectively. As shown in Fig. 1, with increasing h, DeJoin_2 approaches the lower bound of average distortion for various

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2587388, IEEE Transactions on Circuits and Systems for Video Technology

relative payloads, where the relative payload is defined as the ratio of the message length to the number of pixels.

In a similar manner, DeJoin$_2$ can be extended to algorithms for minimizing joint distortion on larger blocks. For instance, the joint distortion defined on 4-pixel blocks can be decomposed into two distortion functions on 2-pixel blocks and thus the message is embedded by executing twice DeJoin$_2$. Recursively, we can get algorithms for minimizing joint distortion on 8-pixel blocks, 16-pixel blocks and so on. In general, the DeJoin method can be applied to pixel blocks with an arbitrary size. For instance, the joint distortion on a 3-pixel block can be decomposed into a distortion function on a 2-pixel block and a distortion function on a single pixel.

The purpose of DeJoin is to reduce the complexity of minimizing joint distortion. When directly extending STCs to k-ary embedding, the computational complexity is $O\left(nk^h\right)$, where n is the number of pixels and h is the height of submatrix used in STCs. Note that k will exponentially increase with the size of the pixel-block, e.g., $k = 9$ for 2-pixel blocks and $k = 81$ for 4-pixel blocks when ± 1 embedding is used. However, by DeJoin, we always execute two-layered STCs with $k = 2$ in n pixels for pixel blocks with any size N by dividing the n pixels into N groups. Therefore, DeJoin can minimize joint distortion with low computational complexity.

IV. Defining Joint Distortion with SMD Principle

In the above section, we propose a general method for embedding messages by minimizing joint distortion. The joint distortion can be defined in several manners, e.g., extending the Gaussian models used in [9] from single pixel to multi-pixels or generalizing the SMD based method [12], [13]. For fairly comparing with the methods in [12], [13], we give some examples on how to define joint distortion based on the principle of SMD.

A. Defining 2-Pixel Joint Distortion

We first define initial distortion on single pixel by using state-of-the-art method for additive distortion, HILL [8], and then define joint distortion on pixel blocks based on the initial distortion. The image is divided into 1×2 non-overlapped blocks. For the block $B_i = (x_{i,1}, x_{i,2})$, we denote the initial distortion on $x_{i,1}$ by $d_1^{(i)}(l)$ for $l \in I$ and the initial distortion on $x_{i,2}$ by $d_2^{(i)}(r)$ for $r \in I$. The joint distortion on B_i is defined by

$$\rho^{(i)}(l, r) = \alpha(l, r) \times \left(d_1^{(i)}(l) + d_2^{(i)}(r)\right). \tag{18}$$

The scaling function $\alpha(l, r)$ is defined in Fig. 2, which shows that changes in the same directions are encouraged by multiplying smaller scaling factors. Messages are embedded by using DeJoin$_2$ for minimizing the joint distortion (18).

$\alpha(l,r)$:

l \ r	-1	0	$+1$
-1	1	2	3
0	2	1	2
$+1$	3	2	1

Fig. 2. Scaling function for joint distortion (18).

We compare the proposed method with HILL [8], MiPOD [10], and UpDist based method [12], [13] for resisting the detection of steganalysis. We denote the method in [12] by CMD (Clustering Modification Directions) and the method in [13] by Synch. In the UpDist based methods [12], [13], an additive scheme is needed to define the initial distortion, in which we also use HILL to define the initial distortion for fair comparison. To reduce complexity of the experiments, we replace double-layered STC with an optimal modification simulation in all methods because double-layered STC can approach the optimal ± 1 embedding.

The detector is trained by using state-of-the-art 34,671-dimensional SRM feature set [18] with the ensemble classifiers [19] with Fisher linear discriminant as the base learner. The performance is evaluated by the ensembles minimal total testing error under equal priors such that

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

where P_{FA} and P_{MD} are the false-alarm probability and the missed-detection probability respectively. The ultimate security is qualified by average error rate $\overline{P_E}$ averaged over 10 random splits of the data set, and larger $\overline{P_E}$ means stronger security. All experiments are conducted on the BOSSbase ver.1.01 image database [20] which contains 10,000 gray-scale images of size 512×512 pixels. As shown in Fig. 3, CMD, Synch, and DeJoin$_2$ can outperform the additive schemes, HILL and MiPod.

With 2-pixel joint distortion (18), DeJoin$_2$ outperforms Synch for various relative payloads, while CMD appears to be slightly more secure than DeJoin$_2$.

B. Enhancing Security by Combining DeJoin with UpDist

Fig. 3 shows that we cannot outperform CMD by defining joint distortion on 2-pixel blocks. In fact, by updating distortion, both CMD and Synch consider the interactive impact of several neighboring pixels, while the joint distortion function (18) only considers the impact of one adjacent pixel.

To further exploit the interactive impact of modifications, we can apply the strategy of UpDist to 2-pixel blocks by taking the 2-pixel block as a super-pixel. To do that, we divide the image into two sub-images after defining 2-pixel joint distortion. Firstly we embed half of payloads into the first sub-image by using DeJoin$_2$. Secondly, we update the joint distortion of the second sub-image according to the changed results of the first sub-image, and then embed the rest payloads into the second sub-image with DeJoin$_2$.

The joint distortion on 1×2 pixel-blocks can reflect the mutual impact of modifications in horizontal directions. To incorporate the mutual impact in vertical directions, we collect all odd rows of the cover image as the first sub-image and all even rows as the second sub-image. The joint distortion on a pixel block in the second sub-image will be updated according to the changed results of the blocks above and under it. For example, we denote the current block in the second sub-image

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2587388, IEEE Transactions on Circuits and Systems for Video Technology

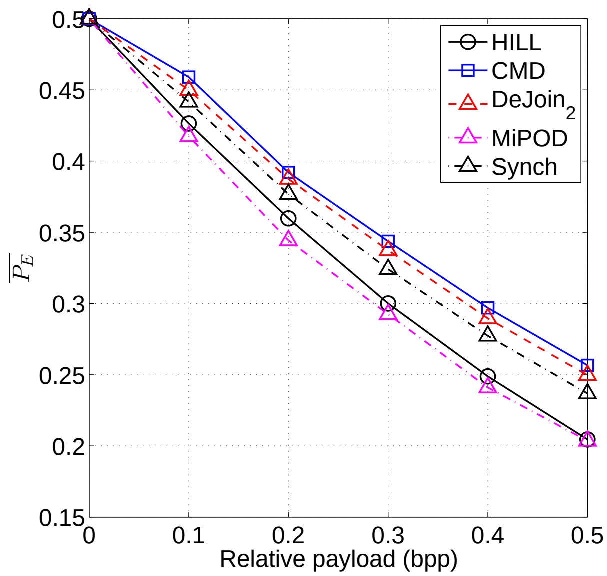


Fig. 3. Comparison between HILL [8], MiPOD [10], Synch [13], CMD [12] and DeJoin$_2$ for resisting the detection of SRM [18].

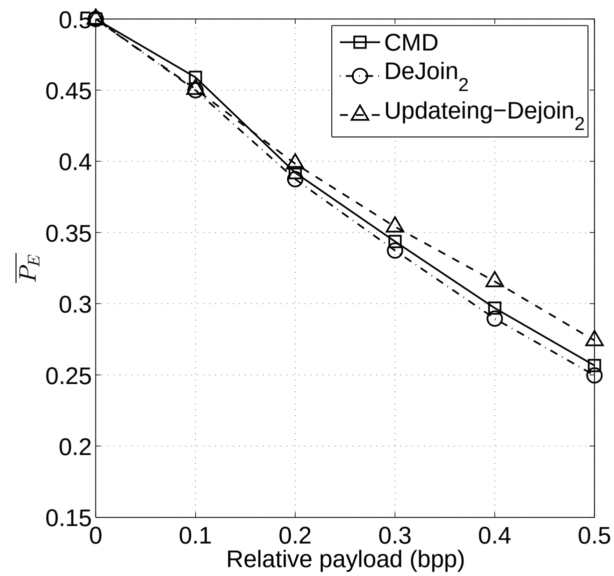


Fig. 4. Comparison between CMD [12], DeJoin$_2$ and Updating-DeJoin$_2$ for resisting the detection of SRM [18].

by (s_1, s_2), the block above it by (a_1, a_2), and the block under it by (u_1, u_2). Assume that in the first round (a_1, a_2) has been changed to $(a_1 + l', a_2 + r')$, and (u_1, u_2) has been changed to $(u_1 + l'', u_2 + r'')$, and then the distortion $\rho(l, r)$ on (s_1, s_2) will be updated to $\rho'(l, r)$ as follows.

$$\rho'(l,r) = \begin{cases} \beta\rho(l,r) & \textit{when } (l,r) = (l',r') \neq (l'',r'') \\ \beta\rho(l,r) & \textit{when } (l,r) = (l'',r'') \neq (l',r') \\ \beta^2\rho(l,r) & \textit{when } (l,r) = (l',r') = (l'',r'') \\ \rho(l,r) & \textit{otherwise} \end{cases} \quad (20)$$

where β $(0 < \beta < 1)$ is a scaling factor which also encourages the modifications in the same directions. If the block (a_1, a_2) is in the top/bottom row, its distortion will be updated only according to the changed results of the block under/above it.

As shown in Fig. 4, by updating 2-pixel joint distortion with setting $\beta = 0.25$ in (20), Updating-DeJoin$_2$ can outperform CMD for relative payloads larger than 0.2.

C. Resisting Selection-Channel-Aware Detection by Using Larger Pixel Blocks

Recent advances in steganalysis show that, by using Selection-Channel-Aware (SCA) features, the Warden can reduce the error rates when detecting adaptive steganography. The maxSRMd2 model [21] is a SCA version of SRM, which has the same dimension with SRM. When using the SCA features, the Warden needs to estimate the change probabilities of each pixel, which is easy for detecting additive schemes, such as HILL and MiPOD.

However, for Synch, CMD, and DeJoin, the Warden can only compute the change probabilities of pixels used in the first round, whose distortion is defined by an additive scheme, and the change probabilities of other pixels may greatly vary in the embedding process. Such variation is dependent on different messages and the stego key. By the stego key, the steganographer can randomly determine which pixels will be used in each embedding round. Therefore, as pointed out by Denemark et al. in [13], the best the Warden can do is to compute change probabilities of all pixels with the additive scheme used in the first round. To fairly compare the performance, we use the same additive scheme, i.e. HILL, to define the basic distortion in CMD, Synch and DeJoin.

The pixels are divided into two subsets in Synch and DeJoin$_2$ while four subsets in CMD. Therefore, the Warden can accurately estimate change probabilities of half of pixels for Synch and DeJoin$_2$ and only quarter of pixels for CMD. That's why CMD performs better under the detection of maxSRMd2 as shown in Fig. 5.

In fact, DeJoin can increase the ability of resisting maxSRMd2 by defining joint distortion in larger pixel blocks. To verify this idea, we design a distortion function on 4-pixel blocks. The image is divided into 2×2 non-overlapped blocks, and we denote each block B_i as

$$B_i = \begin{pmatrix} x_{i1} & x_{i2} \\ x_{i3} & x_{i4} \end{pmatrix} \quad (21)$$

The joint distortion on B_i is defined as follows:

$$\rho^{(i)}(l_1, l_2, l_3, l_4) = \alpha(l_1, l_2, l_3, l_4) \times \left(d_1^{(i)}(l_1) + d_2^{(i)}(l_2) + d_3^{(i)}(l_3) + d_4^{(i)}(l_4)\right), \quad (22)$$

where $(l_1, l_2, l_3, l_4) \in I^4$ and $d_j^{(i)}(l_j)$ denotes the initial distortion on single pixel $x_{i,j}$ defined by HILL for $j \in \{1, 2, 3, 4\}$.

The scaling function $\alpha(l_1, l_2, l_3, l_4)$ also follows the principle of SMD, which is defined as

$$\alpha(l_1, l_2, l_3, l_4) = \frac{1 + o(l_1, l_2) + o(l_1, l_3) + o(l_2, l_4) + o(l_3, l_4)}{1 + s(l_1, l_2) + s(l_1, l_3) + s(l_2, l_4) + s(l_3, l_4)}, \quad (23)$$

where

$$s(l_1, l_2) = \begin{cases} 1 & \text{when } |l_1 - l_2| = 0 \\ 0 & \text{otherwise} \end{cases}, \quad (24)$$

and

$$o(l_1, l_2) = \begin{cases} 1 & \text{when } |l_1 - l_2| = 2 \\ 0 & \text{otherwise} \end{cases}. \quad (25)$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2587388, IEEE Transactions on Circuits and Systems for Video Technology

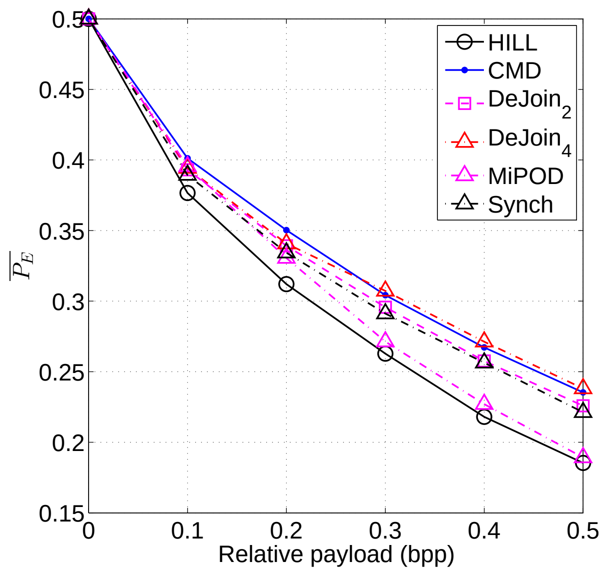


Fig. 5. Comparison between HILL [8], MiPOD [10], Synch [13], CMD [12], DeJoin$_2$ and DeJoin$_4$ for resisting the detection of maxSRMd2 [21].

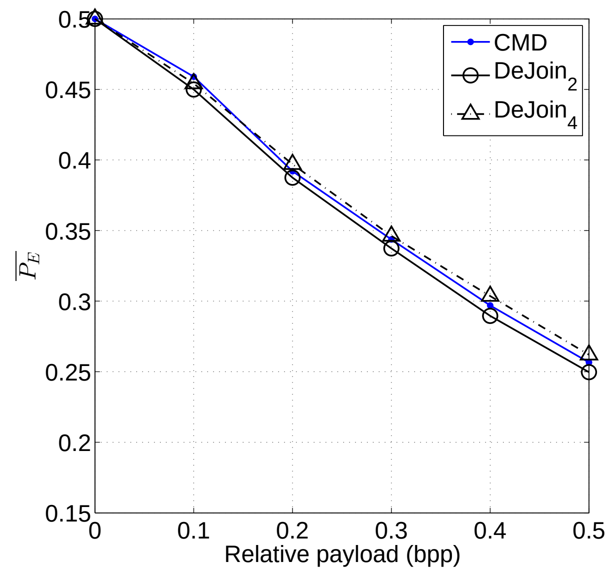


Fig. 6. Comparison between CMD [12], DeJoin$_2$ and DeJoin$_4$ for resisting the detection of SRM [18].

With Eq. (24) and (25), we adjust the embedding cost according to whether two adjacent pixels are modified in the same direction. Therefore the scaling function $\alpha(l_1, l_2, l_3, l_4)$ measures the degree of SMD in a 2×2 pixel-block. We embed messages by using DeJoin$_4$ to minimize the joint distortion defined in (23).

We compare DeJoin$_4$ with other schemes under the detection of maxSRMd2 in Fig. 5, and to give a more detailed presentation, we compare the four SMD-based methods(CMD, Synch, DeJoin$_2$ and DeJoin$_4$) by showing the $\overline{P_E}$ and its standard deviation in Table I. Fig. 5 and Table I show that MiPOD outperforms HILL, and Synch achieves approximately the same level of security with DeJoin$_2$, and DeJoin$_4$ can promote the security of DeJoin$_2$ to reach approximately the same level with CMD. On the other hand, Fig. 6 shows that, by extending the joint distortion to large pixel blocks, DeJoin can also improve the level of security when steganalyzing with SRM.

In order to detect the performance of SMD-based methods above on the images with suppressed noise, we compare four SMD-based methods with payload=0.1bpp on another database BossbaseJQF [22], which is formed by JPEG compressing the images of Bossbase 1.01 with quality factor $QF \in \{75, 80, 85, 90, 95\}$ and then decompressing to the spatial domain. In order to achieve a fair comparison, we also use HILL as the initial distortion for all the SMD-based methods. Fig. 7(a) and Fig. 7(b) show that compared with experiments on Bossbase 1.01, the relative performance of all the schemes keep almost unchanged although the $\overline{P_E}$ here is much lower than the results on the Bossbase 1.01. For example, all the SMD-based methods can promote the security of HILL; DeJoin$_4$ and CMD have very similar performance and can outperform other schemes with different QF.

V. Conclusion

In this paper, we proposed a framework to improve the security of adaptive steganography by defining joint distortion functions on pixel blocks, which exploits the interactive impact of changes between adjacent pixels. A decomposition coding method, DeJoin, is proposed to minimizing the average joint distortion for embedding messages.

The proposed DeJoin method is general, which can be applied to any joint distortion functions on various sizes of pixel blocks, and thus provides a promising tool for minimizing non-additive distortion in steganography. In this paper, we only provide several simple examples about defining joint distortion on small pixel blocks in spatial images. How to define reasonable joint distortion to improve the security of steganography in various kinds of covers will be an interesting direction in the future study.

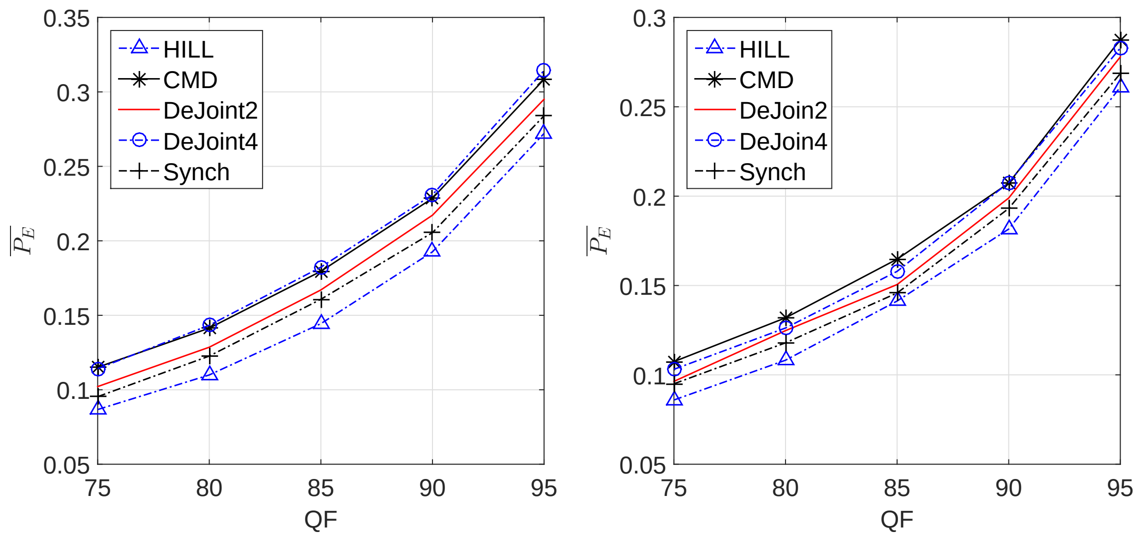
References

[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] Z. Xia, X. Wang, X. Sun, and B. Wang, "Steganalysis of least significant bit matching using multi-order differences," Security and Communication Networks, vol. 7, no. 8, pp. 1283-1291, 2014.
[4] Z. Xia, X. Wang, X. Sun, Q. Liu, and N. Xiong, "Steganalysis of LSB matching using differences between nonadjacent pixels," Multimedia Tools and Applications, vol. 75, no. 4, pp. 1947-1962, 2016.
[5] 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, 2010.
[6] V. Holub and J. Fridrich, "Designing steganographic distortion using directional filters," Proc. of IEEE Workshop on Information Forensic and Security, pp. 234-239, 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, 2013.
[8] B. Li, M. Wang, J. W. Huang, and X. L. Li, "A new cost funtion for spatial image steganpgraphy," Proc. of International Conference on Image Processing, 2014.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2587388, IEEE Transactions on Circuits and Systems for Video Technology

TABLE I
DETECTABILITY IN TERMS OF $\overline{P_E}$ VERSUS EMBEDDED PAYLOAD SIZE IN BITS PER PIXEL (BPP) FOR FOUR SMD-BASED METHODS(BOSSBASE 1.01, MAXSRMD2).

Payload	Synch	CMD	Dejoint2	Dejoint4
0.1	0.38677±0.0024	0.40119±0.0028	0.39303±0.0018	0.39428±0.0021
0.2	0.33132±0.0016	0.35027±0.0029	0.33916±0.0036	0.34072±0.0021
0.3	0.28939±0.0023	0.30417±0.0027	0.29570±0.0031	0.30725±0.0029
0.4	0.25460±0.0030	0.26718±0.0021	0.25758±0.0021	0.27127±0.0021
0.5	0.22373±0.0035	0.23532±0.0023	0.22585±0.0014	0.23809±0.0030



(a) SRM [18]

(b) maxSRMd2 [21]

Fig. 7. Comparison between HILL [8], Synch [13], CMD [12], DeJoin_2 and DeJoin_4 for resisting the detection of SRM [18] and maxSRMd2 [21] on BossbaseJQF.

[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] V. Sedighi, R. Cogranne, J. Fridrich, "Content-adaptive steganography by minimizing statistical detectability," IEEE Trans. on Inf. Forensics Security, vol. 11, no. 2, pp. 221-234, 2016.

[11] 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. 1, pp. 920-935, 2011.

[12] 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, no. 9, pp. 1905-1917, 2015.

[13] T. Denemark and J. Fridrich, "Improving steganographic security by synchronizing the selection shannel," Proc. of 3rd Workshop on I-H&MMSec, Portland, Oregon, June 17-19, 2015.

[14] A. D. Ker, P. Bas, R. Bohme, et. al, "Moving steganography and steganalysis from the laboratory into the real world," in Proc. ftrst Int. Workshop Information Hiding and Multimedia Security, Jun. 17-19, 2013, pp. 45-58.

[15] T. Filler and J. Fridrich, "Gibbs construction in steganography," IEEE Trans. on Inf. Forensics Security, vol. 5, no. 4, pp. 705-720, 2010.

[16] X. Zhang, W. Zhang and S. Wang, "Efficient double-layered steganographic embedding," IET Electronics Letters, vol. 43, no. 8, pp. 482-483, 2007.

[17] W. Zhang, X. Zhang and S. Wang, "A double layered "plus-minus one" data embedding scheme," IEEE Signal Processing Letters, vol.14, no.11, pp. 848-851. Nov. 2007.

[18] J. Fridrich and J. Kodovsky, "Rich models for steganalysis of digital images ," IEEE Trans. on Inf. Forensics Security, vol. 7, pp. 868-882, 2012.

[19] 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, 2012.

[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, 2011.

[21] T. Denemark, V. Sedighi, V. Holub, et al. "Selection-channel-aware rich model for steganalysis of digital images," IEEE National Conference on. Parallel Computing Technologies (PARCOMPTECH), pp. 48-53, 2015.

[22] Sedighi V, Fridrich J, Cogranne R. "Toss that BOSSbase, Alice!" IS&T intl symposium on Electronic Imaging. 2016.