Multimed Tools Appl (2016) 75:281–296
DOI 10.1007/s11042-014-2291-9

Framework for improving the security performance of ordinary distortion functions of JPEG steganography

Fangjun Huang · Hyoung Joong Kim

Received: 25 February 2014 / Revised: 3 August 2014 / Accepted: 19 September 2014 /
Published online: 5 October 2014

Abstract Digital steganography is a new approach for secure communication. Via using it, the sender and the receiver can easily exchange secret message on the Internet without arousing any suspicion. Previously, a lot of ordinary distortion functions for joint photographic experts group (JPEG) steganography have been presented, which can guide the message embedding in the non-zero alternating current (AC) discrete cosine transform (DCT) coefficients of JPEG image. In this paper, we present a framework for improving the security performance of these distortion functions. In our new framework, these ordinary distortion functions are not restricted to evaluating the distortion values of non-zero AC DCT coefficients any more, and their coverage areas will be extended to all DCT coefficients, including the direct current (DC) coefficients and all the zero and non-zero AC coefficients. All the coefficients in JPEG image are divided into two groups: changeable group (CG) and reserve group (RG), respectively. The coefficients that may result in less detectable distortion are grouped into CG and the rest into RG. Via associating the distortion values to coefficients in CG and RG with different strategies, a series of new advanced distortion functions can be generated. The experimental results demonstrate that while applying these advanced distortion functions to JPEG steganography, the statistical characteristics of the carrier image will be preserved better than the prior art, and consequently secure JPEG steganographic schemes can easily be obtained.

Keywords Security · Steganography · Distortion function · JPEG

1 Introduction

Digital steganography is the art and science of passing secret information in a manner that the very existence of hidden message is unknown. Different from traditional watermarking [10, 11], the key concept of steganographic systems is the security performance, i.e., the statistical un-

F. Huang
School of Information Science and Technology, Sun Yat-Sen University, Guangzhou Higher Education Mega Center, Panyu District, Guangzhou 510006, China
e-mail: huangfj@mail.sysu.edu.cn

H. J. Kim (✉)
Graduate School of Information Security, Korea University, Seoul 136-701, South Korea
e-mail: khj-@korea.ac.kr

detectability. It may be influenced by many factors [6], such as the choice of cover object, the type of modification operation on cover elements, the number of embedding changes (related to the payload), and the distortion functions used to identify individual elements of cover that could be modified during embedding. Assuming that the first three factors mentioned above are the same, designing the distortion function would be an important approach to minimizing the impact caused by modification, and thus improve the security performance of steganography.

To minimize the impact caused by data embedding, the sender should choose to modify those elements (pixels/coefficients) in such a way that the caused detectable distortion is as small as possible. Embedding the secret message bits under the guidance of minimizing distortion function can improve the security performance of steganography and has been known for a long time. Previously, Fridrich et al. [4] presented the perturbed quantization (PQ) steganography. As a specific case, they pointed out that the sender can constrain the embedding changes to those DCT coefficients that experience the largest quantization error, i.e., the coefficients with the quantization error of $0.5 \pm \varepsilon$ (ε is a small positive number). Such kind of coefficients, when rounded to the other value, may leave the smallest embedding distortion. In [7], another two adaptive versions of PQ, i.e., texture-adaptive PQ (PQt) and energy-adaptive PQ (PQe) were presented. Through considering the local block content such as texture complexity and energy capacity, JPEG steganography with higher security performance can be obtained. In [2, 31, 32], the authors combined quantization step with quantization error in their distortion function to improve the security performance of JPEG steganography. Besides the quantization step, Wang and Ni [36] presented a new JPEG distortion function with consideration of the entropy of each coefficient block, and the experimental results demonstrate that this entropy based (EB) distortion function may lead to less detectability of steganalyzers. Recently, Huang et al. [13] presented another distortion function for JPEG steganography, which is called new PQ (NPQ). Three factors are considered, i.e., the quantization error, the quantization step and the magnitude of quantized DCT coefficients to be modified. Via nonlinearly combining these three different factors, the new distortion function NPQ can improve the security performance of JPEG steganography significantly.

All the aforementioned ordinary distortion functions are employed to find the non-zero AC DCT coefficients that may result in less detectable distortion for modification. Generally, they are applied together with the utilization of some channel coding techniques, which are called channel-codes embedding in this paper. For example, how to implement PQ distortion function in JPEG steganography was exemplified with the help of wet paper codes [4]. In [19], the authors provided a simple and practical scheme to apply PQ distortion function with utilization of the modified binary Hamming codes. This new embedding strategy allows more than one embedding change in each coefficient block. Via a brute-force search, the modifications are made on those coefficients that may introduce minimal detectable distortion, and thus improving the security performance of the corresponding JPEG steganography. According to the number of allowable changing bits in each coefficient block, these modified matrix encoding (MME) schemes are called MME2, MME3, etc. Similar approach can also be made based on BCH (Bose, Chaudhuri and Hocquenghem) codes [26] to improve the efficiency of channel-codes embedding as described in [32, 39]. However, since the decoding of BCH codes is much more complicated than Hamming codes, some specific techniques need to be adopted by the sender to reduce the time complexity and storage complexity in the embedding process. Recently, Filler et al. [2] provided the syndrome-trellis codes (STCs), which can be utilized for embedding while minimizing an arbitrary additive distortion function with a performance near the theoretical bound. This new methodology can directly improve the security performance of many existing steganographic schemes, allowing them to communicate larger payloads at the same embedding distortion or to decrease the distortion for a given payload.

Note that with the rapid development of steganography, the counterpart steganalysis has also made much achievement in the past few years. Via employing the techniques in the fields such as signal processing and pattern recognition, the existence of the hidden material may be detected with a high accuracy rate. For some classical steganographic schemes such as least significant bit (LSB) substitution, the cover and stego images can easily be discriminated [27]. Moreover, not only the message length can be estimated accurately [1], but also the embedded message can be located [17] under some special conditions. What is worth mentioning, a lot of efficient universal steganalyzers have been proposed in the past few years, such as moments based steganalyzers [38], Markov process based steganalyzers [35], MM (merging Markov and DCT features) [28], CC-PEV (cartesian-calibrated Pevný) [20], SPAM (subtractive pixel adjacency matrix) [29], CDF (cross domain feature) [21], NJ (neighboring joint density-based JPEG steganalyzer) [25], and rich models [23, 24]. Though these universal steganalyzers can only be applied in laboratory conditions as pointed by Ker et al. [18], e.g., the steganalyst needs to have the perfect knowledge of the cover source and the embedding algorithm, these steganalyzers still pose a great threat to the security of today's steganographic schemes.

In this paper, we present a new framework for improving the security performance of those previously proposed ordinary distortion functions. Firstly, the ordinary distortion functions' applied range is extended from non-zero AC DCT coefficients to all DCT coefficients, thus all DCT coefficients are utilized for channel-codes embedding and the efficiency (i.e., the number of bits embedded per embedding change [5]) can be improved. Secondly, in order to minimize the detectable distortion that may be introduced in the embedding process, all DCT coefficients are divided into two groups: changeable group (CG) and reserve group (RG). Note that in our framework, any coefficient can be modified in the embedding process and no coefficient is considered as un-changeable. The coefficients that may result in less detectable distortion are grouped into CG and the rest are into RG. Several general rules for dividing the coefficients into CG and RG are given and one scenario for dividing the coefficients into CG and RG is exemplified, which can be utilized to form a series of new advanced distortion functions. Experimental results demonstrate that while applying these advanced distortion functions to JPEG steganography with channel-codes embedding techniques, JPEG steganographic schemes with higher security performance can easily be obtained.

The rest of this paper is organized as follows. In Section 2, the proposed new framework is introduced. Experimental results and analysis are illustrated in Section 3, and the conclusion is drawn in Section 4.

2 Proposed framework

Suppose the raw, uncompressed side-image is available to the sender, which is called pre-cover image as that in [16]. The DCT coefficients that have been divided by quantization steps and not yet rounded are called un-rounded DCT coefficients, and those that have been divided by quantization steps and rounded are called quantized DCT coefficients, respectively. The previously proposed ordinary distortion functions such as PQ [4], NPQ [13] and EB [36] will be drawn into our framework to generate a series of new advanced distortion functions. To make this paper self-contained, we will introduce the PQ, NPQ and EB distortion functions firstly.

2.1 Previously proposed ordinary distortion functions

Without loss of generality, the quantized and un-rounded DCT coefficients utilized for data hiding are represented by $C=(c_1,c_2,\ldots,c_N)$ and $C'=(c'_1,c'_2,\ldots,c'_N)$, respectively, where N

represents the number of coefficients in the quantized and un-rounded DCT coefficient sequence. The relationship between $c_i(1 \le i \le N)$ and $c'_i(1 \le i \le N)$ is as follows.

$$c_i = round\left(c'_i\right) \tag{1}$$

where *round* (x) is a function that rounds the element x to its nearest integer. Note that in Eq. 1, c_i represents the quantized DCT coefficient that is obtained in JPEG compression without secret message embedding. Suppose that while embedding the secret message the modification needs to be made on c_i, and the coefficient after being modified is represented by s_i.

2.1.1 PQ distortion function

PQ distortion function is represented as follows.

$$d_{c_i}^{PQ} = \left|\left|c_i - c'_i\right| - \left|s_i - c'_i\right|\right| \tag{2}$$

where $|x|$ is a function that returns the absolute value of the corresponding element x. For any coefficient c_i, the PQ distortion value $d_{c_i}^{PQ}$ can be computed according to Eq. 2. As pointed out in [4, 19], while embedding the secret message bits, the sender should select those coefficients with minimal PQ distortion values for modification.

2.1.2 NPQ distortion function

NPQ can be regarded as an improved version of PQ with considering the quantization step and the magnitude of the quantized DCT coefficient to be modified. Suppose the quantization step associated with the coefficient c_i is q_i. According to [13], NPQ distortion function is represented as follows.

$$d_{c_i}^{NPQ} = d_{c_i}^{PQ} \times (q_i)^{\lambda_1} / (\mu + |c_i|)^{\lambda_2} \tag{3}$$

where λ_1 and λ_2 are two parameters that are used to control the impacts caused by q_i and $|c_i|$, respectively. As recommended in [13], the two control parameters λ_1 and λ_2 can be selected in the rage of (0, 1]. The parameter μ is utilized to avoid the zero divisors in Eq. 3. When NPQ is only utilized to compute the distortion value corresponding to the non-zero AC DCT coefficients, μ is selected as 0. Otherwise, the parameter μ can be selected as a small number, e.g., the number 1. For any coefficient c_i, the NPQ distortion value $d_{c_i}^{NPQ}$ can be computed according to Eq. 3. As pointed out in [13], while embedding the secret message, the sender should select those coefficients with minimal NPQ distortion values for modification.

2.1.3 EB distortion function

As pointed out in [36], texture regions in the image may have larger entropy than those in smooth regions, thus the entropy can be employed to evaluate the texture complexity of the image regions and construct the distortion profile. With considering the entropy of the 8×8 block of the non-zero quantized AC DCT coefficients, the EB distortion function is represented as follows.

$$d_{c_i}^{EB} = \left(\frac{q_i(|s_i - c'_i| - 0.5)}{H\left(B^{(c_i)}\right)}\right)^2 \tag{4}$$

where $B^{(c_i)}$ represents 8×8 quantized DCT coefficient block that includes the coefficient c_i. Assume there are K different non-zero quantized AC DCT coefficients in block $B^{(c_i)}$, the entropy of the corresponding coefficient block is defined as $H\left(B^{(c_i)}\right) = -\sum_{k=1}^{K} h_k^{B^{(c_i)}} \log h_k^{B^{(c_i)}}$, where $h_k^{B^{(c_i)}}$ $(k = 1, 2, \ldots, K)$ represent the probabilities of all non-zero quantized AC DCT coefficients in block $B^{(c_i)}$. Note that if K=0, a small value is assigned to $H\left(B^{(c_i)}\right)$ directly. Wang and Ni [36] applied the EB distortion function on all the non-zero quantized AC DCT coefficients, and good experimental results have been obtained.

2.2 The proposed framework

As mentioned above, in [13, 19, 36], the proposed ordinary distortion functions PQ, NPQ and EB are only utilized to guide the modification on those non-zero AC DCT coefficients. In our new framework, we will extend their coverage range to all the DCT coefficients. That is, all the DCT coefficients will be utilized for channel-codes embedding. One of the immediate benefits of utilizing all DCT coefficients (including numerous zero coefficients) for channel-codes embedding is that the embedding efficiency (i.e., the number of bits embedded per embedding change [5]) can be improved greatly. For example, when MME2 or MME3 codes is utilized for channel-codes embedding, k secret message bits can be embedded into the corresponding DCT coefficient block with length of 2^k-1 by making at most two or three embedding changes. As seen, with a larger value of k, the channel-codes embedding will be conducted more efficiently. Note that the value of k is determined by the length (denoted by L) of all secret message bits to be embedded and the number (denoted by N) of DCT coefficients that can be utilized for data hiding. Generally, the maximum value of k that satisfies $\frac{k}{2^k-1} \geq \frac{L}{N}$ is selected in the embedding process. As seen, with the increasing of N, a larger value of k can be obtained. Note that utilizing some zero coefficients in low frequencies for data hiding can not only improve the channel-codes embedding efficiency, but also improve the security performance of JPEG steganography, which have been demonstrated in [12, 15].

However, utilizing all DCT coefficients for channel-codes embedding may simultaneously introduce a threat to the security of the corresponding steganography, since some coefficients that may introduce much detectable distortion may be modified inevitably in the embedding process. Thus in our new framework, we will divide all DCT coefficients into two groups: CG and RG. The coefficients that may introduce less detectable distortion are divided into CG, and the rest coefficients are divided into RG. Via associating distortion values to the coefficients in CG and RG with different strategies, the modifications can mainly be made on those coefficients in CG in the embedding process. That is, on one hand, the channel-codes embedding can be conducted more efficiently; on the other hand, the modification can be made on those coefficients that may introduce minimal detectable distortion as much as possible.

2.2.1 How to divide coefficients into CG and RG

The statistics of DCT coefficients are complicated and they may interact with each other. Moreover, the statistics of DCT coefficients may also have a very close relationship with the secret message bits to be embedded, and the type of embedding operation that modifies the coefficients, etc. There is no absolute standard for determining which coefficient should be divided into CG or RG. That is, it is not easy for us to derive an optimal strategy for dividing the coefficients into CG and RG in our framework. However, a series of suboptimal scenarios

can be found easily as pointed out in [14]. Here are some general rules for dividing coefficients into CG and RG.

Figure 1 illustrates the standard JPEG quantization table corresponding to the quality factor (QF) of 75, which is popularly used by JPEG images on the internet. Generally, the magnitudes of the elements in the quantization table are increased according to the *zig-zag* scanning order. As we know, the inverse block DCT is a linear transformation. Thus, the modification on the quantized DCT coefficient associated with smaller quantization step may result in less distortion in spatial domain. Based on this observation, in JPEG steganography, the modification should be made on those coefficients belonging to the relatively low frequencies as much as possible. Consequently, in our framework, most of the low frequency coefficients (including the numerous zero coefficients) are divided into CG, whereas the rest high frequency coefficients are divided into RG.

Let it be noted that in JPEG image the 8×8 coefficient blocks are associated with the same quantization table, and modification on coefficients in the same frequency may result in the same distortion in spatial domain. However, it does not mean that the same number of quantized DCT coefficients can be selected for modification from each 8×8 block. In Fig. 2, two 8×8 quantized DCT coefficient blocks of "lena.jpg" image with QF=75 are illustrated. Figure 2a corresponds to a texture or noisy region while Fig. 2b corresponds to a smooth region. As seen, the distribution of low frequency coefficients in Fig. 2a is much more difficult to model than that in Fig. 2b. That is, modifications made on the low frequency coefficients in Fig. 2a may result in low statistical detectability, whereas modifications on low frequency coefficients in Fig. 2b may easily be captured by today's universal steganalyzers. Based on this observation, the number of coefficients selected for modification should be determined according to the texture complexity of each 8×8 block of the carrier image. Thus in our framework, in the texture blocks more coefficients are divided into CG, whereas in the smooth blocks fewer coefficients are divided into CG.

As we mentioned before, the coefficients in any frequency can be modified in the embedding process as if the modification is less enough to a limited extent. In [14], we demonstrated that all the AC coefficients are considered as changeable coefficients, and good security performance could still be obtained. However, this scenario is not recommended in this paper, since it may introduce some weakness in the corresponding steganographic schemes. In Fig. 3, we illustrate the ratios of zero coefficients in each frequency of some JPEG images with QF=75. Figure 3a corresponds to the ratios of "lena.jpg", and Fig. 3b

$$\begin{bmatrix} 8 & 6 & 5 & 8 & 12 & 20 & 26 & 31 \\ 6 & 6 & 7 & 10 & 13 & 29 & 30 & 28 \\ 7 & 7 & 8 & 12 & 20 & 29 & 35 & 28 \\ 7 & 9 & 11 & 15 & 26 & 44 & 40 & 31 \\ 9 & 11 & 19 & 28 & 34 & 55 & 52 & 39 \\ 12 & 18 & 28 & 32 & 41 & 52 & 57 & 46 \\ 25 & 32 & 39 & 44 & 52 & 61 & 60 & 51 \\ 36 & 46 & 48 & 49 & 56 & 50 & 52 & 50 \end{bmatrix}$$

Fig. 1 Standard JPEG quantization table corresponding to QF=75

$$\begin{bmatrix} -38 & -9 & 13 & 0 & -1 & 1 & 0 & 0 \\ 16 & 10 & -2 & -1 & 2 & 0 & 0 & 0 \\ -8 & -2 & 1 & -1 & 2 & 0 & 0 & 0 \\ -7 & 5 & 1 & 2 & 0 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 & -1 & 0 & 0 & 0 \\ -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad \begin{bmatrix} 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

(a) (b)

Fig. 2 Two 8×8 DCT coefficient blocks of "lena.jpg" with QF=75. **a** The 2008th (from top to bottom and then left to right) 8×8 block. **b** The 2373th (from top to bottom and then left to right) 8×8 block

corresponds to the average ratios of all 10,000 images in BOSSBase image dataset [3]. The horizontal axis represents the index of frequency according to *zig-zag* scanning order and the vertical axis represents the ratio of zero coefficients. It is easy to find out that in most of the high frequencies, the ratio of zero coefficients is near 100%. Note that if we compute the distortion values according the aforementioned ordinary distortion functions directly, those high frequency coefficients may be associated with a very small distortion value. For example, in Eqs. 2, 3 and 4, if the round errors of some un-rounded coefficients in high frequency are approximately equal 0.5, the corresponding distortion values will be about zero and they may be selected for modification in the embedding process. These modifications should be avoid as far as possible since they may be captured by today's feature based steganalyzers or some specific JPEG steganalyzers.

In addition, As pointed in [9], the modification on DC coefficients may introduce some cyclical distortion and thus lower the security performance of the corresponding JPEG steganography. In order to make our framework more generally applicable, the DC coefficients are excluded from the changeable coefficient in our framework. In the following, one scenario for dividing the coefficients into CG and RG will be exemplified.

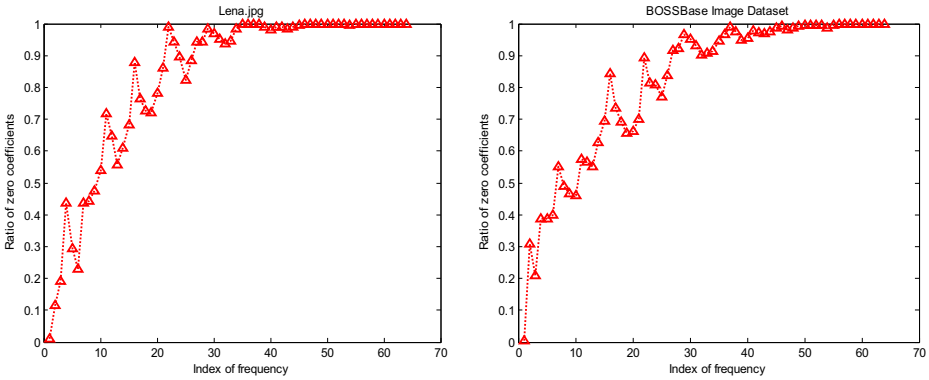

Fig. 3 The ratios of zero coefficients in different frequencies **a** "lena.jpg" with QF=75. **b** BOSSBase images with QF=75

2.2.2 Exemplified scenario for dividing the coefficients into CG and RG

In our exemplified scenario, the quantized DCT coefficients are divided into CG and RG according to the statistics of pre-cover image in spatial domain. Suppose the standard deviation of pixel values in each 8×8 block of the pre-cover image is $P_i(1\le i\le N)$, where N represents the total number of 8×8 blocks in the pre-cover image. The average value of all the standard deviations is $\overline{P}=\frac{1}{N}\sum_{i=1}^{N}P_i$, and the maximum value among all the standard deviations is $P_{\max}=\max(P_1,P_2,\ldots,P_N)$. In each block, the number of AC DCT coefficients that belongs to CG is computed as follows.

$$B_i=\begin{cases}1, & \text{if } 0\le P_i<\frac{1}{32}P\\ \left\lfloor\frac{1}{2}\times 64\times\left(P_i/P\right)\right\rfloor, & \text{if } \frac{1}{32}P\le P_i<P\\ \left\lfloor\frac{1}{2}\times 64\times\left(1+P_i/P_{\max}\right)\right\rfloor, & \text{if } P\le P_i<P_{\max}\\ 63 & \text{if } P_i=P_{\max}\end{cases} \tag{5}$$

where $B_i(1\le i\le N)$ represents the number of AC DCT coefficients that should be divided in CG in each 8×8 block, and $\lfloor x\rfloor$ is a function that rounds the element x to its nearest integer less than or equal to x. In Eq. 5, the number 64 represents that there are 64 DCT coefficients in each 8×8 block. In this scenario, the $B_i(1\le i\le N)$ changeable coefficients in each block are selected according to the *zig-zag* scanning order, and the rest AC and DC coefficients are considered as reserve coefficients. Other methods for dividing the coefficients into CG and RG may still work, e.g., we can change the number 32 to 31 or 30 in Eq. 5, and the obtained distortion function may still work. Here, we only try to illustrate the applicability of our framework and do not try to make a clear boundary between CG and RG.

2.2.3 Proposed advanced distortion function

In our framework, the impact caused by the modifications of coefficients in CG and RG can still be measured using some ordinary distortion functions firstly. Then, those obtained distortion values associated with the coefficients in RG are multiplied by a penalty factor, which is a big value. The proposed advanced distortion function is defined in Eq. 6.

$$d_{c_i}^{ADV}=d_{c_i}^{ORD}\times(1+\delta) \tag{6}$$

In Eq. 6, the $d_{c_i}^{ORD}$ represents the impact caused by modification operation on coefficient c_i, which is computed according to the ordinary (abbreviated as "*ORD*") distortion functions such as PQ, NPQ and EB. The penalty factor δ is selected as a big value (e.g., 10^6) for the coefficient $c_i\in RG$, otherwise it is selected as 0. According to Eq. 6, for any coefficient c_i in the input image, the advanced (abbreviated as "*ADV*") distortion value $d_{c_i}^{ADV}$ can be easily computed.

As seen, the distortion values associated with the coefficients in CG may be much less than that in RG in general in our advanced distortion function. When embedding message bits with some channel-codes embedding strategy as in [2, 13, 19, 32, 36, 39], several alternative solutions may be produced and those coefficients in CG that may result in less detectable distortion will take precedence for modification. On the other hand, even if all the alternative

solutions are restricted to those coefficients in RG, the coefficients in RG associated with smaller ordinary distortion values will still take precedence for modification. That is, the advanced distortion functions can pilot us to make as less distortion as possible in the embedding process, and thus the security performance of JPEG steganography can be improved. Note that as we mentioned before, no coefficient is considered as un-changeable in our framework, and any DCT coefficient can be modified if needed in the embedding process. That is also the main reason why we call these coefficients changeable and reserve coefficients, not changeable and un-changeable coefficients.

2.2.4 Modification way on the DCT coefficients

Via applying our advanced distortion functions to JPEG steganography, no special processing needs to be made on those quantized DCT coefficients with values of +1 and −1 as that in [13, 19, 36]. Note that in [13, 19, 36], the applied range of ordinary distortion functions is the non-zero quantized AC DCT coefficients. If the coefficient with value of +1 or −1 is flipped to 0, the recipient will not be able to accurately locate the corresponding non-zero coefficients utilized for channel-codes embedding in the transmitting end, and the embedded secret message bits may not be extracted successfully. Thus, special modification operation should be made by the sender on those quantized coefficients with values of +1 and −1. For example, in [13, 19, 36] the quantized coefficients with values of +1 and −1 can only be flipped to +2 and −2, respectively.

Since the advanced distortion functions generated from our framework are applied on all the DCT coefficients, i.e., all the coefficients are utilized for channel-codes embedding, no such special modification operation needs to be made while applying our advanced distortion functions to JPEG steganography. For any coefficient $c_i(1 \le i \le N)$ to be modified, the operation is conducted as follows.

$$s_i = \begin{cases} c_i + 1, & \text{if } \left(c_i - c_i'\right) \le 0 \\ c_i - 1, & \text{if } \left(c_i - c_i'\right) > 0 \end{cases} \tag{7}$$

where s_i is the coefficient after having been modified.

A special note of interest is that while applying those advanced distortion functions generated from our framework to JPEG steganography, the sender should first divide all the DCT coefficients into CG and RG. However, the sender does not need to share the dividing scenario with the recipient, since they (i.e., the sender and recipient) both use all the DCT coefficients to conduct channel-codes embedding. The recipient does not need to locate the DCT coefficients in CG or RG in the receiving end, and he/she can exchange the secret message with the sender easily via selecting the same channel-codes embedding strategy.

According to our above description, the main difference between our new advanced distortion function and those ordinary distortion function is the coverage area which is extended from non-zero AC DCT coefficients to all DCT coefficients. While applying these new advanced distortion functions to JPEG steganography, the channel-codes embedding strategy remains the same and the modification way on the DCT coefficients is even more simple. Though the number of coefficients need to be processed is increased via using our advanced distortion functions, the computational complexity does not change. Thus our advanced distortion functions can be applied to JPEG steganography conveniently as those ordinary distortion functions.

3 Experimental results

In this section, experimental results and analysis are presented to demonstrate the effectiveness of our proposed framework. The test image set consists of 10,000 pre-cover images which are downloaded from the BOSSBase image dataset [3]. All the images are with the size of 512× 512. In the following, the JPEG compressed image without any message embedding is called cover image. The cover and stego images are created using the same JPEG encoder as that in [33], and the quality factor is selected as 75 in all of our testing. The secret message bits are randomly generated, and the embedding rates are represented in terms of *bpac* (bits per non-zero quantized AC DCT coefficients) values.

In our experiments, the PQ, NPQ and EB are selected as the ordinary distortion functions for demonstrating the effectiveness of our framework, and their corresponding advanced distortion functions are represented as AdvPQ, AdvNPQ and AdvEB, respectively. We have applied the aforementioned three ordinary distortion functions and their corresponding advanced distortion functions to JPEG steganography with three different channel-codes embedding strategies (i.e., MME2, MME3 and STC) for a detailed comparison. Note that the two control parameters of NPQ are selected as (λ_1=0.5,λ_2=0.2), and the height of pseudorandom sub-matrix of STC is selected as H=10 in all our testing.

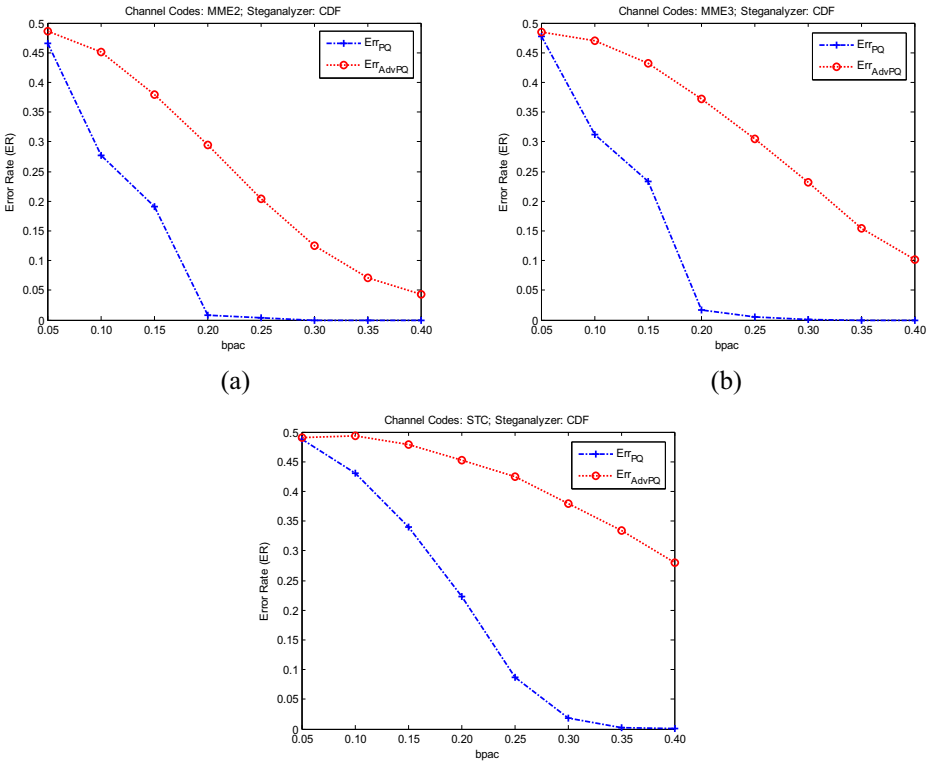

Fig. 4 The detection error rates corresponding to PQ and AdvPQ with different channel-codes embedding strategies. **a** MME2 **b** MME3 **c** STC

The security performance of our framework is tested with one of the most popular JPEG steganalyzers CDF [21]. It is a combined version of CC-PEV and SPAM. The 548-dimensional CC-PEV feature vector is mainly extracted from JPEG domain and the 686-dimensional SPAM feature vector is extracted from spatial domain. Through combing the CC-PEV and SPAM feature vectors, we can get 1,234-dimensional CDF feature vector. The feature vector or its improved versions are popularly utilized in evaluating the security performance of some classical algorithms such as F5 [37] and MB1 [33], and a lot of modern steganographic schemes [8, 9, 30, 34].

The ensemble classifier presented in [22] is employed in our testing with default parameters (publicly available implementation of the ensemble classifier can be downloaded from http://dde.binghamton.edu/download/ensemble). It is a fully automatic framework with an efficient utilization of out-of-bag (OOB) error estimates for stopping criterion. As pointed out in [22], the proposed ensemble classifier consists of a lot of base learners independently trained on a set of cover and stego images. The decision threshold of each base learner is adjusted to minimize the total detection error under equal priors on the training set:

$$P_E = \min_{P_{FA}} \frac{1}{2}\left(P_{FA} + P_{MD}(P_{FA})\right) \tag{8}$$

where P_{FA} , P_{MD} are the probabilities of false alarm and missed detection, respectively.

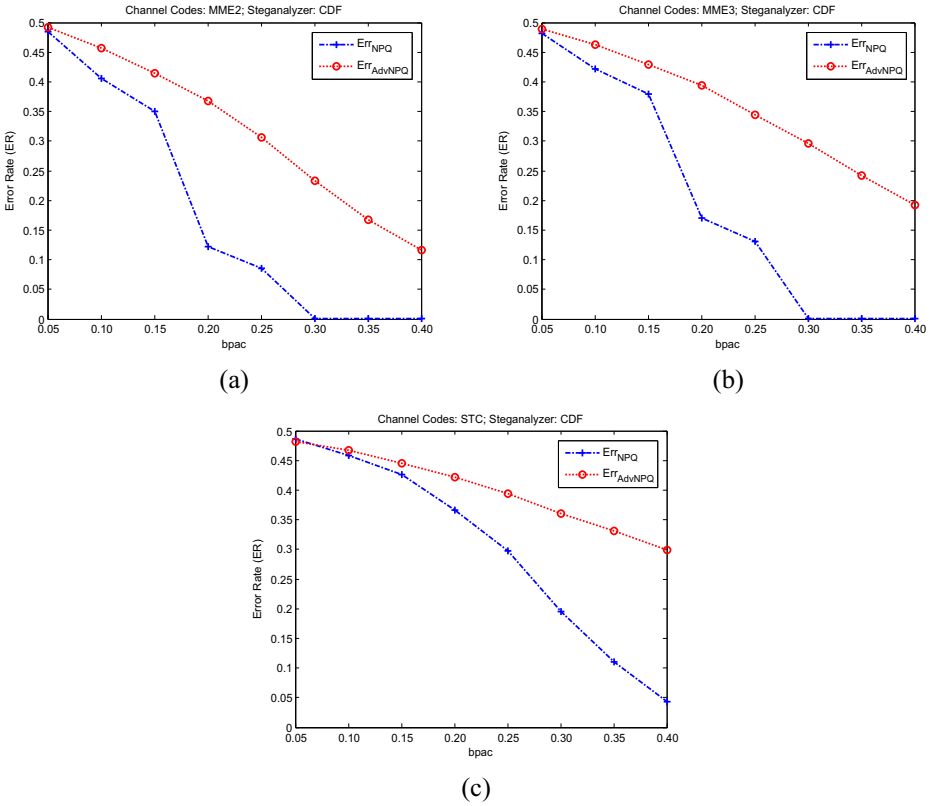

Fig. 5 The detection error rates corresponding to NPQ and AdvNPQ with different channel-codes embedding strategies. **a** MME2 **b** MME3 **c** STC

The detection error rates corresponding to different ordinary and advanced distortion functions (i.e., PQ and AdvPQ, NPQ and AdvNPQ, EB and AdvEB) are illustrated in Figs. 4, 5 and 6, respectively. In these figures, the horizontal axes represent the *bpac* values, and the vertical axes represent the detection error rates. For the aforementioned eighteen steganographic schemes (there are 6 different distortion functions and each distortion function has been tested with MME2, MME3 and STC three different channel-codes embedding strategies), the test embedding rates are increased from 0.05 *bpac* to 0.40 *bpac* with the step size of 0.05. The selected channel-codes embedding strategy and the steganalyzer are illustrated in the title of each figure. For the aforementioned ordinary and advanced distortion functions, their corresponding detection error rates are represented as Err_{PQ} , Err_{AdvPQ} , Er_{NPQ} , Err_{AdvNPQ} , Err_{EB} and Err_{AdvEB} , respectively.

It is observed from Figs. 4, 5 and 6 that for any ordinary distortion function, with incorporating it into our proposed framework, the security performance of the resulted JPEG steganographic scheme can be greatly improved. At some circumstances, the final detection error rates will increase 35 percentage points or even more, e.g., PQ distortion function with MME3 embedding strategy at the embedding rate of 0.20 *bpac*, and EB distortion function with MME3 embedding strategy at the embedding rate of 0.30 *bpac*. Our experimental results also demonstrate that the selections of

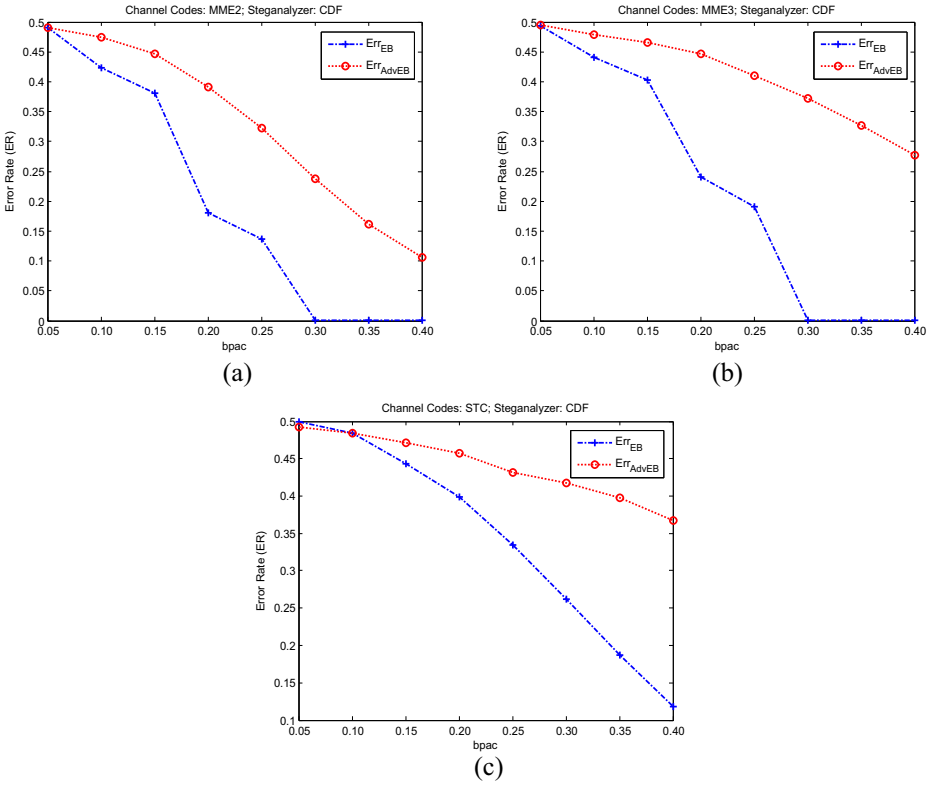

Fig. 6 The detection error rates corresponding to EB and AdvEB with different channel-codes embedding strategies. **a** MME2 **b** MME3 **c** STC

Table 1 The detection error rates corresponding to different the number of modified coefficients

The number of modified coefficients	1,000		2000		3,000	
	CG	RG	CG	RG	CG	RG
Error rates (%)	13.20	0.02	5.46	0	2.83	0
Visual Quality (PSNR)	46.2	40.6	43.2	37.6	41.5	35.8

ordinary distortion function and the embedding codes are also two important factors that may influence the final security performance of the obtained steganographic schemes. That is, with selecting more efficient embedding strategy or more optimized ordinary distortion function, the effectiveness of our proposed framework may be improved further. For more detail, please refer to Figs. 4, 5 and 6.

At last, we also conduct some experiments to demonstrate the effectiveness of our exemplified scenario for dividing the coefficients into CG and RG. There are three steps in our test: 1) randomly select 1,000–3,000 coefficients belonging to CG or RG for modification (i.e., randomly minus or plus one); 2) use CDF to extract features from the cover and modified images; 3) employ the ensemble classifier to differentiate the cover and modified images. The experimental results are shown in Table 1. As seen, when the coefficients belonging to RG are selected for modification, the cover and modified images are easier to be differentiated. For example, even if the number of coefficients (belonging to RG) selected for modification is as few as 1,000 bits, the final differentiation error rate is about 0.02 %. However, if the 1,000 coefficients selected for modification belongs to CG, the final differentiation error rate will be as high as 13.2 %. Note that dividing the coefficients into CG and RG can not only improve the security performance of the JPEG steganography, it can also improve the visual quality of the stego images. In Table 1, we also illustrate the average peak signal-to-noise ratios (PSNR) between the cover images and the modified images. It is observed from Table 1 that when modifications are made on those coefficients belonging to CG, much better visual quality can be obtained.

4 Conclusions

JPEG is one of the most common image formats produced by digital cameras, scanners, and various photographic image capture devices nowadays. Therefore, hiding secret message into JPEG images may provide effective camouflage. In order to improve the security performance of JPEG steganography, generally there are two approaches, i.e., channel-codes embedding and distortion function designing. However, different from that of channel-codes embedding there is a theoretical bound that can be utilized to judge the embedding efficiency, while for distortion function designing, there is no absolute standard until now. Designing distortion function of JPEG steganography is a hard and open problem. In this paper, we presented a framework for designing distortion functions of JPEG image with pre-cover image. The main contributions of our framework are as follows.

1) The proposed framework has good practicability. Via using it, a lot of advanced distortion functions can easily be generated to form a series of JPEG steganographic schemes with

high security performance. It is an efficient way to confuse today's steganalyzers and move steganography from the laboratory into the real world.

2) Our proposed framework is an open system. It will not be constrained to the aforementioned dividing scenario and ordinary distortion functions. Other dividing scenarios and ordinary distortion functions can be adopted easily in our framework to generate a series of new advanced distortion functions.
3) Several general rules for evaluating the detectable distortion associated with the DCT coefficients are illustrated in this paper, which may provide some insight into the kind of distortion function designing work in the future.

Acknowledgements This work was partially supported by the National Natural Science Foundation of China (61173147), the Korea Foundation for Advanced Studies' International Scholar Exchange Fellowship for the academic year of 2013-2014, the Fundamental Research Funds for Central Universities (12lgpy31), and the Project Sponsored by the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry ([2012]1707).

References

1. Dumitrescu S, Wu X, Wang Z (2003) Detection of LSB steganography via sample pair analysis. IEEE Trans Signal Process 51(7):1995–2007
2. Filler T, Judas J, Fridrich J (2010) Minimizing additive distortion in steganography using Syndrome-Trellis Codes. IEEE Trans Inform Forensic Secur 6(3):920–935
3. Filler T, Pevny T, and Bas P (2010) BOSS (Break Our Steganography System). http://www.agents.cz/boss, July 2010
4. Fridrich J, Goljan M, and Soukal D (2004) Perturbed quantization steganography with wet paper codes, In Proceedings of the ACM Workshop on Multimedia & Security, Magdeburg, Germany, September 20–21, pp. 4–15.
5. Fridrich J, Soukal D (2006) Matrix embedding for large payloads. IEEE Trans Inform Forensic Secur 1(3): 390–395
6. Fridrich J, Lisoněk P, Soukal D (2007) On Steganographic embedding efficiency. In Proceedings of 8th Information Hiding Workshop. Lect Notes Comput Sci 4437:282–296
7. Fridrich J, Pevný T, and Kodovský J (2007) Statistically undetectable JPEG steganography: dead ends, challenges, and opportunities. In Proceedings of the ACM Workshop on Multimedia and Security, Dallas, Texas, September 20–21, pp. 3–14.
8. Holub V and Fridrich J (2012) Designing steganographic distortion using directional filters. In Proceedings of 4th IEEE International Workshop on Information Forensics and Security, Tenerife, Spain, December 2–5.
9. Holub V and Fridrich J (2013) Digital image steganography using universal distortion. In Proceedings of 1th ACM Workshop on Information Hiding and Multimedia Security, Montpellier, France, June 17–19.
10. Horng S.-J, Farfoura ME, Fan P, Wang X, Li T and Guo J.-M (2013) A low cost fragile watermarking scheme in H.264/AVC compressed domain. Multimedia Tools and Applications: 1–27.
11. Huang H-C, Pan J-S, Huang Y-H, Wang F-H, Huang K-C (2007) Progressive Watermarking Techniques Using Genetic Algorithms. Circ Syst Sig Process 26(5):671–687
12. Huang F, Shi YQ, Huang J (2010) New JPEG steganographic scheme with high security performance. In Proceedings of 9th International Workshop on Digital Watermarking. Lect Notes Comput Sci 6526:189–201
13. Huang F, Huang J, Shi YQ (2012) New channel selection rule for JPEG steganography. IEEE Trans Inform Forensic Secur 7(4):1181–1191
14. Huang F, Luo W, Huang J, and Shi YQ (2013) Distortion function designing for JPEG steganography with uncompressed side-image. In Proceedings of 1th ACM Workshop on Information Hiding and Multimedia Security, Montpellier, France, June 17–19.

15. Huang F, Kim HJ, and Zhang D (2013) Efficiency of Frequency Selection in JPEG Steganography. In Proceedings of 3th International Conference on Multimedia Technology, Guangzhou, China, Nov. 29 - Dec. 1.
16. Ker AD (2007) A fusion of maximal likelihood and structural steganalysis. In Proceedings of 9th Information Hiding Workshop. Lect Notes Comput Sci 4567:204–219
17. Ker AD (2008) Locating steganographic payload via WS residuals. In Proceedings of 10th ACM Workshop on Multimedia and Security, Oxford, UK. Sep. 22–23, pp. 27–32.
18. Ker AD, Bas P, Böhme R, Cogranne R, Craver S, Filler T, Fridrich J and Pevný T (2013) Moving steganography and steganalysis from the laboratory into the real world. In Proceedings of 1th ACM Workshop on Information Hiding and Multimedia Security, Montpellier, France, June 17–19.
19. Kim Y, Duric Z, Richards D (2007) Modified matrix encoding technique for minimal distortion steganography. In Proceedings of 8th Information Hiding Workshop. Lect Notes Comput Sci 4437: 314–327
20. Kodovský J, and Fridrich J (2009) Calibration revisited. In Proceedings of the ACM Multimedia & Security Workshop, Princeton, New Jersey, September 7–9, pp. 63–74.
21. Kodovský J, Pevný T, and Fridrich J (2010) Modern steganalysis can detect YASS. In Proceedings of SPIE, Electronic Imaging, Security Forensics of Multimedia XII, San Jose, California, Jan. 17–21, vol. 7541, pp. 0201–0211
22. Kodovský J, Fridrich J, Holub V (2012) Ensemble calssifiers for steganalysis of digital media. IEEE Trans Inform Forensic Secur 2(7):432–444
23. Kodovský J, Fridrich J (2012) Steganalysis of JPEG images using rich models. In Proceedings of SPIE, Electronic Imaging, Media Watermarking, Security, and Forensics XIV, vol. 8303, San Francisco, CA, January 22–26, pp. 0A 1–13.
24. Li F, Zhang X, Chen B, Feng G (2013) JPEG steganalysis with high-dimensional features and Bayesian ensemble classifier. IEEE Sig Process Lett 20(3):233–236
25. Liu Q, Sung A, Qiao M (2011) Neighboring joint density-based JPEG Steganalysis. ACM Trans Intell Syst Technol 2(2):1–16
26. Moon TK (2005) (2005) Error Correction Coding, Mathematical Methods and Algorithms. Wiley, Hoboken
27. Patsakis C, Aroukatos N (2014) LSB and DCT Steganographic Detection Using Compressive Sensing. J Inf Hiding Multimedia Sig Process 5(1):20–32
28. Pevný T, and Fridrich J (2007) Merging Markov and DCT features for multi-class JPEG steganalysis. In Proceedings of SPIE Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents IX, San Jose, California, Jan. 28 - Feb. 1, vol. 6505, pp. 03.1-03.13
29. Pevný T, Bas P, Fridrich J (2010) Steganalysis by subtractive pixel adjacency matrix. IEEE Trans Inform Forensic Secur 52(2):215–224
30. Pevný T, Filler T, Bas P (2010) Using high-dimensional image models to perform highly undetectable steganography. In Proceedings of 12th Information Hiding Workshop. Lect Notes Comput Sci 6387:161–177
31. Sachnev V, Kim HJ, and Zhang R (2009) Less detectable JPEG steganography method based on heuristic optimization and BCH syndrome coding. In Proceedings of the ACM Workshop on Multimedia & Security, Princeton, New Jersey, Sep. 7–9, pp. 131–140.
32. Sachnev V, Kim HJ (2012) Modified BCH data hiding scheme for JPEG steganography. EURASIP J Adv Signal Process 2012:89–98
33. Sallee P (2005) Model based methods for steganography and steganalysis. Int J Image Graph 5(1):167–190
34. Solanki K, Sarkar A, Manjunath BS (2007) YASS: Yet another steganographic scheme that resists blind steganalysis. In Proceedings of 9th Information Hiding Workshop. Lect Notes Comput Sci 4567:16–31
35. Shi YQ, Chen C, Chen W (2007) A Markov process based approach to effective attacking JPEG steganography. In Proceedings of 8th International Workshop on Information Hiding. Lect Notes Comput Sci 4437: 249–264
36. Wang C, Ni J (2012) An efficient JPEG steganographic scheme based on block-entropy of DCT coefficients. In Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Kyoto, Japan, Mar. 25–30, pp. 1785–1788.
37. Westfeld A (2001) High capacity despite better steganalysis (F5-a steganographic algorithm). In Proceedings of 4th Information Hiding Workshop. Lect Notes Comput Sci 2137:289–302
38. Xuan G, Shi YQ, Gao J, Zou D, Yang C, Zhang Z, Chai P, Chen CH, Chen W (2005) Steganalysis based on multiple features formed by statistical moments of wavelet characteristic functions. In Proceedings of 7th Information Hiding Workshop. Lect Notes Comput Sci 3727:262–277
39. Zhang R, Sachnev V, Kim HJ (2009) Fast BCH syndrome coding for steganography. In Proceedings of 11th Information Hiding Workshop. Lect Notes Comput Sci 5806:48–58

Fangjun Huang received his B.S. degree from Nanjing University of Science and Technology, China, in 1995, his M.S. degree and Ph.D. degree from Huazhong University of Science and Technology, China, in 2002 and 2005, respectively.

From June of 2009 to June of 2010, he was a post-doctoral researcher in Department of Electrical and Computer Engineering, New Jersey Institute of Technology, New Jersey, USA. He is now an associate professor with the School of Information Science and Technology, Sun Yat-Sen University, Guangzhou, China. He is also a visiting scholar from August of 2013 to August of 2014 in Korea University, Seoul, Korea. His research interests include steganography, steganalysis, and digital forensics.

Hyoung Joong Kim got bachelor's degree in electrical engineering in 1978 from Seoul National University, and he received his master and Ph.D. degrees in control and instrumentation engineering in 1986 and 1989, respectively, from Seoul National University. He was a visiting scholar from 1992 to 1993 at University of Southern California, CA, USA.

He was a Director of International Association of Cryptologic Research (IACR). He was a guest editor of several international journals such as IEEE Transactions on Circuits and Systems for Video Technology. He proposed many suggestions to MPEG and made them accepted as International Standard (IS). He is currently a professor of Korea University, Seoul, Korea. He is also the President of Society of Digital Contents and President of Smart Media Association.