A FRAMEWORK OF ENHANCING IMAGE STEGANOGRAPHY WITH PICTURE QUALITY OPTIMIZATION AND ANTI-STEGANALYSIS

Guo-Shiang Lin\(^1\), Yi-Ting Chang\(^2\), and Wen-Nung Lie\(^2\)

\(^1\)Dept. of Computer Science and Information Engineering, Da-Yeh University, Chang-Hua, Taiwan, R.O.C.
Address: 112 Shan-Jiau Rd., Da-Tsuen, Changhua County 515, Taiwan, R.O.C.
Tel: 886-4-8511888 ext. 2405
Fax: 886-4-5811350
Email: khlin@mail.dyu.edu.tw

\(^2\)Dept. of Electrical Engineering, National Chung Cheng University, Chia-Yi, Taiwan, R.O.C.
Address: 168, University Rd., Min-Hsiung Chia-Yi 621, Taiwan, R.O.C.
Tel: 886-5-2720411 ext. 33211
Fax: 886-5-272-0862
Email: wnlie@ee.ccu.edu.tw

ABSTRACT

Picture quality and statistical undetectability are two key issues related to steganography techniques.

In this paper, we propose a closed-loop iterative computing framework that enhances image steganographic schemes with picture quality optimization and anti-steganalysis. To achieve our goal, an anti-steganalysis tester and a controller based on a simulated annealing algorithm are incorporated into the loop to guarantee an accurate convergence of searches. A cost function is defined to integrate three performance indices, namely, the mean square error, the HVS deviation, and the differences in statistical features. The proposed scheme embeds data in the quantized DCT domain; hence, it is applicable to both raw and JPEG images. Compared with its prior version of non-optimized algorithm, a better image quality (with an average PSNR improvement of 6.68 dB) can be achieved and, at the same time, the anti-steganalysis property is significantly enhanced.

Index Terms— Steganography, steganalysis, human visual system, data embedding
I. INTRODUCTION

Imperceptibility is clearly the most important requirement in steganographic schemes [1]. That is, the modifications between the cover medium and its stego version should be slight and transparent to a human eye. However, nowadays, even slight modifications might be discoverable by using an adequate mechanism with the aid of a computer, e.g., steganalysis [18], [19], [20], [21], [33]–[35]. Generally, the main weakness of most steganographic schemes is that changes in an image’s statistics after data embedding may be detected by steganalyzers.

Currently, most steganalytic schemes take advantage of significant changes in an image’s statistical characteristics to distinguish a stego media from the original one. According to the operating domain, steganalytic schemes can be categorized into three classes: spatial, transform, and hybrid. Some spatial-domain steganalytic algorithms [18], [19], [20], [21], [33] have been developed to detect certain steganographic schemes. For example, Fridrich et al. [20] proposed a steganalysis technique based on the observation that bit-planes in typical images are more or less correlated so that the LSB (least significant bit) plane can be estimated from the other seven planes. However, the estimation becomes less reliable as the content of the LSB plane is randomized further. Kong et al. [21] proposed evaluating an image’s complexity by using a statistical filter to determine whether it contains secret messages. This approach is based on a phenomenon that randomness of the LSB plane’s content becomes heavier after information is hidden. In [33], the correlation between higher bits of pairs of adjacent pixels is explored to detect the presence of data embedding based on a multiple-LSB replacement. Lin et al. [38] proposed a steganalytic scheme against the MBNS (multiple-base notational system) based on a phenomenon that small remainders often occur for any given base in a stego image.

Some steganalytic methods [22]–[25] have been proposed in the transform domain. For example, Chandramouli [23] modeled common steganographic schemes as a linear transform between the cover and the stego images, which can be estimated if at least two copies of the stego images (for the same cover) can be obtained. This problem is similar to a blind source separation problem, which can be solved by using the
independent component analysis (ICA) [26] technique. In [24], Harmsen et al. proposed a steganalytic scheme for dealing with information hiding schemes that mix a secret and a cover signal in an addition rule. The phenomenon whereby the center of mass of the histogram characteristic function of a stego image moves to the left of (or remains the same as) that of the cover image is observed and exploited to distinguish stego images from plain images. In [25], Lyu et al. use first- and higher-order magnitude and phase statistics as features after QMF (Quadrature Mirror Filters) decomposition of an image. A detector is then formed by inputting these features to an SVM (Support Vector Machine) classifier. In [34], the joint statistics of DCT coefficients between a JPEG image and its cropped version are used as input features to SVM to identify the adopted steganographic algorithm from among 6 ones.

Some hybrid schemes [2], [3] have also been developed. Avcibas et al. [2] note that quality degradation after smoothing or low-pass filtering is different for images with and without embedded information. In other words, by observing the difference in quality between a test image and its smoothed version, it is possible to identify the existence of hidden messages. Based on this concept, Avcibas et al. proposed using regression analysis with several quality measuring operators for steganalysis. Lie et al. [3] proposed a steganalytic scheme where two features are computed and classified with a neural classifier. The features are the gradient energy, which characterizes spatial-domain variations of the graylevels between adjacent pixels, and the statistical variance of the Laplacian parameters used to measure the distribution of spectral coefficients in local macroblock areas. This proposed steganalyzer does not target a specific steganographic scheme. In [35], temporal correlation is analyzed to achieve video steganalysis.

In recent years, some researchers [13], [14], [17], [27], [28] have emphasized an additional requirement for statistical undetectability when embedding information, so as to develop a scheme that is capable of resisting steganalysis. This technique, still in its infancy, is called “anti-steganalysis.”

Basically, there are two approaches for enhancing the capability of anti-steganalysis: non-parametric methods and parametric methods. “Non-parametric” means that anti-steganalysis is achieved not by tuning the parameters in the embedding procedure. For example, Wu et al. [13] propose a scheme in which a genetic
algorithm (GA) is used to iteratively modify pixel graylevels such that the difference between the statistical properties of the cover image and its stego version does not exceed a certain tolerance. The process continues until a steganalytic scheme in the loop fails. However, Wu et al.’s steganographic scheme can not guarantee a zero BER (Bit Error Rate) after data extraction. Moreover, the scheme operates on individual 8×8 pixel blocks, not a full frame, implying a possible failure to resist steganalysis that considers the statistical features of more than one block. The methods proposed in [27] and [28] construct a stego image by replacing each block of a cover image with another block that has similar statistical properties. Since altering pixel values or finding similar blocks, under the constraint of remaining statistically undetectable, is somewhat difficult, it is expected that time complexity of these methods [13], [27], [28] will be large.

“Parametric” means that anti-steganalysis can be achieved by tuning the parameters used in the embedding procedure directly, e.g., reducing the magnitude of the inserted noise, reducing the embedding rate, or changing the embedding positions [17]. Compared with non-parametric methods, parametric approaches seem more efficient, but they may experience problems with the overheads generated by recording parameters or with adapting to extracting embedded data.

The above-mentioned limitations motivate us to develop a multi-objective data embedding architecture that can satisfy the fundamental requirements of steganography simultaneously, e.g., high picture quality, low extraction BER, and anti-steganalysis. In addition, to consider the proposed framework’s applicability to existing embedding schemes, we adopt a parametric approach that can construct an embedding architecture with an anti-steganalysis property. Therefore, an SA (Simulated Annealing)-based algorithm is developed in the proposed framework to augment the compressed-domain steganographic scheme in [5]. The proposed SA-based algorithm modifies the transform coefficients so that some performance indices (e.g., MSE, HVS (human visual system) deviation, and the differences between statistical features) are optimized subject to certain constraints (e.g., anti-steganalysis).

II. SYSTEM ARCHITECTURE
According to [1] and [28], the following factors should be considered when designing a steganographic scheme: (1) imperceptibility, (2) statistical undetectability, (3) embedding capacity, and (4) BER after data extraction. Note that, since it is important to transmit a secret message faithfully, BER should be guaranteed zero in case the stego image is free from attacks. This is not true for schemes in [9], [13], [30], and [32], but nevertheless satisfied by the scheme adopted here. To develop compressed-domain embedding schemes, another factor, file size variation, should also be taken into account [4]. However, some of the above factors may conflict with one another. For example, increasing the embedding capacity might reduce the imperceptibility and increase the statistical detectability. Visual artifacts may occur even if secret bits are embedded into inadequate positions in the compressed domains. Since it is difficult to arrange a tradeoff between them by using an open-loop steganography architecture, we propose a closed-loop architecture that optimizes multiple objectives simultaneously.

To arrange a tradeoff between multi-objectives, we need some performance indices for system evaluation. In our system, three performance indices are evaluated and used to guide the modification of pixel values or transform coefficients for embedding. The indices are: (1) \( f_1 \): MSE (mean square error), (2) \( f_2 \): HVS deviation, and (3) \( f_3 \): anti-steganalysis (in terms of the differences in statistical features).

In [4] and [28], MSE and Watson’s metric are used to measure picture quality. A lower MSE indicates a higher PSNR and a good Watson’s metric indicates a stego image that is indistinguishable from the original version. In the proposed architecture, we also use two performance indices, MSE and HVS deviation, to measure imperceptibility. Though embedding a message of smaller size (i.e., lower embedding capacity) can make stego images nearly indistinguishable from their cover images, it is less meaningful (we have to optimally increase the embedding rate without loss of statistical undetectability). As mentioned in the previous section, steganalysis [2], [3] is often based on the statistical features computed from the spatial or transform domain. Thus, a steganographic scheme that can resist steganalysis has to consider variations in the statistical features of these two domains. To this end, a steganalytic subsystem from our prior work [3], which analyzes the gradient energy (in the spatial domain) and the Laplacian parameter (in the DCT domain) as statistical
features, is incorporated into the evaluation loop. In principle, anti-steganalysis can be further enhanced as more features examined in other steganalysis systems [2], [20]-[25], [33]-[35] are integrated into this proposed framework.

Most steganalytic schemes determine whether an image contains confidential messages by analyzing the image’s statistical properties. In contrast to steganalysis, anti-steganalysis tries to make statistical properties of a stego image similar to those of its original version [11], [15]. Herein, the performance index $f_1$ is utilized to measure the differences between the statistical properties of the stego and the cover image.

To integrate the three indices into the evaluation loop, we define a cost function $E$ as follows:

$$E = w_1 \times f_1 + w_2 \times f_2 + w_3 \times f_3,$$  

where $w_1$, $w_2$, and $w_3$ are predefined weights. Each index $f_i$ ($i=1$, 2, 3) is first normalized before it can be used in Eq. (1):

$$\hat{f}_i = \frac{(f_i - f_{i,\min})}{(f_{i,\max} - f_{i,\min})},$$  

where $f_i$ and $\hat{f}_i$, $i=1$, 2, 3, denote the original index value and its normalized value, respectively, and $f_{i,\min}$ and $f_{i,\max}$ represent the minimum and maximum values of $f_i$, respectively. Figure 1 shows the architecture of our proposed steganographic system. Note that anti-steganalysis is not only used as a hard constraint, but its analyzed features (i.e., the gradient energy and the Laplacian parameter) are also included in the cost function (as $f_3$) for optimization. A case of $w_3 = 0$ is worthy of notice. It implies that the embedding scheme is blind to possible steganalyzers and anti-steganalysis of a stego image can be achieved only in a trial-and-error manner. This situation is called “blind anti-steganalysis” and will be implemented in experiments that follow.

There is currently no systematic way to decide/optimize the weights $w_1$, $w_2$, and $w_3$. One criterion to determining them is: the more general a factor is, the larger its corresponding weight is. Another criterion is: user’s preference or importance of the performance index. The optimization is then performed based on the given set of weightings.
However, it is difficult to achieve anti-steganalysis and optimize the cost function in a one-shot manner, especially in the compressed domain. Hence, an iterative procedure is needed. To achieve convergence efficiently, some optimization search methods, e.g., GA, SA, and particle swarm optimization, can be used. In our model, the optimization search is conducted by the SA mechanism [6], [36], with good initial solutions estimated by some analytical techniques.

![Diagram](image)

**Fig. 1** The proposed closed-loop architecture for data embedding with enhanced anti-steganalysis

**III. THE PROPOSED STEGANOGRAPHIC SCHEME**

**A. Basic steganographic scheme**

For simulating our closed-loop framework, we adopt the steganographic scheme proposed in [5] as a base and enhance it with the proposed computing architecture. The scheme in [5] is based on a compressed-domain, implying that data/messages are embedded in the quantized DCT coefficients.

First, a cover image is divided into several embedding units (EUs), each of which is composed of one or more 8×8 pixel blocks. An integer $L$ is determined before data embedding, which defines the embedding capacity of each EU as $\log_2 L$ bits. It is assumed that $L$ is known to both the transmitter and the receiver.

Let a secret message be $M = \{m_n | n=1, 2, \ldots, N_m\}$. We would like to hide the message element $m_n$ in the $n$-th EU with $0 \leq m_n < L$. The embedding procedure is repeated $N_m$ times to embed $M$ in the whole image. For raw cover images, $M$ can be embedded after they have been Discrete-Cosine-Transformed (DCT, for each partitioned block of 8×8 pixels) and the DCT coefficients have been quantized (according to the quantization parameter or quality factor). For popular JPEG images, the message $M$ can be embedded after the bit-stream
has been Variable-Length-Decoded (VLD) and the decoded symbols have been inversely zigzag ordered. For both cases, it is assumed that the embedding output is in JPEG format. To deal with JPEG images efficiently, we use the quantization index (QI) of each DCT coefficient, i.e., the above-mentioned decoded symbol (without de-quantization), to hide data. This strategy prevents quantization attacks on the embedded data. The steps of the embedding procedure are as follows.

H1. Obtain the DCT QIs of the given JPEG image after VLD processing. (For raw images, transform each partitioned 8×8 pixel block and quantize each resulting DCT coefficient to obtain its QI)

H2. For each EU, calculate the sum, $S$, of the DCT QIs (except the DC component) and its remainder with respect to $L$ as

$$S = \sum_{(k,l)} x_{i,j}^q(k,l)$$

and

$$r = S \mod L, \quad r \geq 0$$

where $X_{i,j}^q = \{x_{i,j}^q(k,l)\}_{0 \leq k,l \leq 7}$ is the DCT QI representation of the $(i,j)$-th block, $(k,l)$ is the spectral position index, and mod represents the modulo operator. Obviously, we have the remainder $0 \leq r < L$.

H3. Compute the modification $d^o$ for data embedding as

$$d_1 = |m - r|, \quad d_2 = L - d_1,$$

and

$$d^o = \min(d_1, d_2).$$

It is clear that $0 \leq d^o \leq \lfloor L/2 \rfloor$.

H4. Modify the DCT QIs in the $n$-th EU such that the following rules are satisfied:
\[
\tilde{S} = S - d^o, \quad \text{if } r > m_n \text{ and } d_1 < d_2
\]
\[
\tilde{S} = S + d^o, \quad \text{if } r > m_n \text{ and } d_1 \geq d_2
\]
\[
\tilde{S} = S + d^o, \quad \text{if } r < m_n \text{ and } d_1 < d_2
\]
\[
\tilde{S} = S - d^o, \quad \text{if } r < m_n \text{ and } d_1 \geq d_2
\]
\[
\tilde{S} = S, \quad \text{if } r = m_n
\]

where \(\tilde{S}\) denotes the modified version of \(S\). Note that the basic scheme does not constrain the manner in which DCT QI’s are selected and modified. The detailed design will be discussed later.

H5. Repeat Steps H1–H4 for all EUs.

H6. Perform VLC (variable length encoding) to create a stego JPEG image.

Equations (5), (6), and (7) show that, as \(L\) increases (i.e., the embedding capacity becomes larger), the allowable range of \(d^o\) is also increased; hence it is expected that the picture quality might be degraded.

The steps for extracting the hidden message from the quantized DCT domain at the receiver side are described as follows [5].

E1. Decode the received JPEG image to obtain the DCT QIs of each block.

E2. Calculate \(\hat{S}\) for each EU as

\[
\hat{S} = \sum_{(k,l) \neq (0,0)} \hat{X}_{ij}^q(k,l),
\]

where \(\hat{X}_{ij}^q = \{\hat{S}_{ij}(k,l) | 0 \leq k, l \leq 7\}\) is the DCT QI representation of the \((i,j)\)-th block for the decoded JPEG image.

E3. Compute \(\hat{r} = \hat{S} \mod L\). Then, \(\hat{r}\) represents the retrieved message element \(\hat{m}_n\).

E4. Repeat Steps E1–E3 for all EUs to get all the message data.

According to the above descriptions, the steganographic scheme is blind and hidden messages can be extracted without knowing the embedding positions (i.e., the DCT coefficients selected for embedding).

B. Measuring HVS deviation
As described in Section II, our system evaluates HVS deviation for data embedding. This is achieved by setting a JND (Just Noticeable Difference) mask for each DCT QI and measuring the deviation. Following [7], [8], [10], [11], [14], [30], [31], we consider three characteristics of the HVS model when calculating the JND mask, namely, frequency sensitivity, luminance sensitivity, and contrast masking.

Frequency sensitivity describes the human eye’s sensitivity to sine waves with various frequencies. The JND mask resulting from frequency sensitivity is image-independent and often indicated by the default JPEG quantization table [7], [30]. Luminance sensitivity measures the detectability threshold of noise in a constant background. Considering the above two factors, the JND mask can be expressed as follows:

\[
M_{(i,j)}^L(k,l) = T(k,l)\left(\frac{x_{(i,j)}^0(0,0)}{\bar{x}^o(0,0)}\right)^{\omega_L},
\]

where \( T(k,l) \) denotes the frequency sensitivity; \( x_{(i,j)}^0(0,0) \) and \( \bar{x}^o(0,0) \) denote the DC component of the \((i,j)\)-th block and the average of all DC components in an image, respectively; and \( \omega_L \) is a parameter used to adjust the degree of luminance sensitivity. The value of \( \bar{x}^o(0,0) \) is 1024 for an 8-bit grayscale representation and the default value of \( \omega_L \) is 0.649 [10], [31]. Equation (9) shows that the luminance sensitivity increases as the background gets brighter.

Contrast masking refers to the effect of reducing the visibility of one signal in the presence of another signal. When contrast masking is taken into account, the JND mask can be expressed as follows [11], [31]:

\[
M_{(i,j)}^C(k,l) = \max\left\{M_{(i,j)}^L(k,l), x_{(i,j)}^0(0,0)^{\omega_C}\left[M_{(i,j)}^L(k,l)\right]^{-\omega_C}\right\},
\]

where \( \omega_C \) is an exponent between 0 and 1. Typical values of \( \omega_C \) are 0.7 for AC coefficients and 0 for the DC component.

C. The proposed closed-loop embedding scheme

Next, we summarize the proposed data embedding procedures, which are based on the closed-loop architecture illustrated in Fig. 1. Only non-zero DCT QIs are selected for data embedding.
MH1. Obtain the DCT QIs of the given JPEG image after the VLD process. (For raw images, transform each partitioned 8×8 pixel block and quantize each resulting DCT coefficient to get its QI)

MH2. For each EU, calculate $S$ and $r$ by Eqs. (3) and (4), respectively.

MH3. Compute $d''$ by Eqs. (5) and (6).

MH4. Optimally modify non-zero DCT QI’s in the $n$-th EU with

$$\delta = \{ \delta_{(i,j)}(k,l) \mid \delta_{(i,j)}(k,l) \geq 0 \text{ and } x_{(i,j)}^q(k,l) \neq 0 \}$$

to denote the vector of modifications on QIs for each EU subject to Eq. (7). We discuss the optimization method in detail in Section IV.

MH5. Evaluate Eq. (1). The three indices in Eq. (1) are calculated as follows:

$$f_1 = \frac{1}{H \cdot W} \sum_{m=1}^{H} \sum_{n=1}^{W} \left( I_{(i,j)}^c(m,n) - I_{(i,j)}^s(m,n) \right)^2,$$

$$f_2 = \sum_{(k,j) \in 0} \left( \delta_{(i,j)}(k,l) - M_{(i,j)}^C(k,l) \right) \cdot U \left( \delta_{(i,j)}(k,l) - M_{(i,j)}^C(k,l) \right),$$

$$f_3 = \left| f_3^{1,s} - f_3^{1,o} \right| + \left| f_3^{2,s} - f_3^{2,o} \right|,$$

where $I_{(i,j)}^c(m,n)$ and $I_{(i,j)}^s(m,n)$ denote the $(m,n)$-th pixel value in the $(i,j)$-th EU of the cover and stego images, respectively; $H$ and $W$ denote the EU height and EU width, respectively; $U(\cdot)$ denotes the unit step function ($U(x) = 1$ for $x > 0$ and $U(x) = 0$ for $x \leq 0$); and $f_3^{i,s}$ and $f_3^{i,o}$ $(i=1,2)$ denote two statistical features [3] measured from the considered EUs of the stego and the cover images, respectively. Note that all three indices are measured for an EU under processing.

Eq. (11) is equivalent to calculating MSE. Eq. (12) measures the amount of HVS/JND deviation from the mask/threshold for non-zero DCT QI’s. Note that modifications are only made to avoid breaking “zero” runs. The advantage of maintaining zero runs is that it results in the least variation in the size of the JPEG file after data embedding. Eq. (13) measures the difference between two statistical features of a cover image and its stego version in the current iteration. Note that both the performance index $f_3$ and the anti-steganalysis test are required to ensure that the stego image is resistant to
steganalysis at a given embedding capacity.

MH6. Repeat Steps MH1–MH5 for all EUs to create a stego image.

D. Complexity analysis

We now consider how to alter non-zero DCT QIs with $\delta$ in Step MH4 such that $|\tilde{S} - S| = d^o$ (i.e., Eq. (7)) is satisfied. For example, $d^o$ can be evenly distributed between all non-zero DCT QIs, or it can be concentrated in only one non-zero DCT QI. The combination for selecting DCT QIs for data embedding is

$$
\sum_{i=1}^{d^o} \binom{N_o}{i},
$$

where $N_o$ denotes the number of non-zero DCT QIs in the considered EU. For $N_o = 100$ and $d^o = 10$, the number of combinations when $d^o$ is evenly distributed among 10 DCT QIs is more than $2.5E+151$. It is expected that there will be a large number of possible solutions to alter DCT QIs for data embedding in an EU whose $N_o$ is large. In addition, as the HVS/JND constraint is considered when adjusting the modification $\delta_{(i,j)}(k,l)$ of each chosen DCT QI (hence, $d^o$ is not evenly distributed), the number of combinations increases further. Therefore, we need an effective search method to find a solution $\delta$ that minimizes the cost function in Eq. (1). The strategy proposed in [5] modifies DCT QI’s in a priority order, which sorts the quantization error in a decreasing manner. The approach only uses the PSNR index to guide the data embedding process and no optimization procedure is involved.

However, our goal is to improve the steganographic scheme in [5] by properly distributing $d^o$ among non-zero DCT QI’s such that the multi-objective cost function $E$ in Eq. (1) is minimized and the anti-steganalysis property is strengthened. As mentioned earlier, the computational complexity of finding optimal solutions by exhaustive search is extremely high. Here, a more efficient method based on the SA algorithm is adopted instead.

IV. SA-BASED OPTIMIZATION
Simulated annealing (SA) algorithm is one kind of optimization method [6], [36]. Similar to genetic algorithm (GA), SA looks for the minimum of an objective function (cost function) that represents the goodness of a complicated system. Since SA is composed of downhill iteration and controlled uphill steps, it is possible to flee from local minima. Here we do not discuss the SA algorithm in detail due to space limitations. Readers may refer to related literature [6], [36]. In short, three important elements of the SA algorithm should be considered: the initial solution, neighboring solutions, and the cost function. The difference between the SA algorithm and the steepest descent algorithm is the Metropolis rule, which allows the temporary selection of a neighboring solution with a higher cost function, but it eventually achieves the minimum (i.e., a capability of escaping from a trap) [6], [36]. Clearly, the cost function defined in Eq. (1) is highly related to the operation of the SA algorithm and the three performance indices are tuned by the SA algorithm.

A. Initial solution

A good initial solution results in a faster convergence [6], [36]. In our model, an analytic solution based on the minimization of MSE \((f_1)\), subject to the HVS model \((f_2)\), is calculated and used as the initial solution.

Let \(d = \tilde{S} - S\). According to Eq. (7), \(d\) can be expressed as

\[
d = d^* \cdot \text{sgn}(r - m_e) \cdot \text{sgn}(d_1 - d_2) .
\]

(14)

Then, the embedding rule in the proposed scheme can be formulated as follows:

\[
x_{(i,j)}^h(k,l) = x_{(i,j)}^g(k,l) + \text{sgn}(d) \cdot \delta_{(i,j)}(k,l),
\]

(15)

and

\[
\sum_{(k,l) \neq (0,0)} x_{(i,j)}^h(k,l) = \tilde{S},
\]

(16)

where \(X_{ij}^h = \{x_{(i,j)}^h(k,l) \mid 0 \leq k,l \leq 7\}\) is the stego version and \(\text{sgn}(\cdot)\) is the sign function. Recall that our scheme decomposes the difference \(|d|\) into \(\delta(k,l)\)'s, i.e., \(\sum_{(k,l)} \delta(k,l) = |d|\). On the other hand, the degree of modification \(\delta(k,l)\) is subject to the mask value \(M^C(k,l)\). Hence, it is possible that an EU may not have
enough capacity to embed the message $m_n$. For these two cases (sufficient/insufficient capacity), we propose different methods for finding initial solutions.

**Case 1: Sufficient capacity**

According to [4], the norm of distortion will be preserved in both the image and the DCT domains. Hence, minimizing the distortion in either domain should achieve the same result. The problem of minimizing MSE for each EU, denoted as $\varepsilon_{EU}$, subject to the HVS model, can be expressed as

$$
\min \left\{ \varepsilon_{EU} = \sum_{(k,j)} \left( \left| x_{(i,j)}^h(k,l) - x_{(i,j)}^q(k,l) \right| \cdot \Delta(k,l) \right)^2 \right\}
$$

subject to

$$
\sum_{(k,j)} \left| x_{(i,j)}^h(k,l) - x_{(i,j)}^q(k,l) \right| \cdot \Delta(k,l) \leq \sum M_{(i,j)}^C(k,l)
$$

where $\Delta(k,l)$ is the quantization step size for the $(k,l)$-th coefficient; and $M_{(i,j)}^C(k,l)$ is the mask value subject to the HVS model, as calculated in Eq. (10). Note that, to simplify the problem of finding initial solutions, we have combined the HVS constraints (i.e., $\left| x_{(i,j)}^h(k,l) - x_{(i,j)}^q(k,l) \right| \cdot \Delta(k,l) \leq M_{(i,j)}^C(k,l)$) of the DCT coefficients. The constrained optimization problem in Eq. (17) can be converted into an unconstrained optimization problem by the Lagrangian multiplier method. Then, Eq. (17) can be expressed as

$$
\min \left\{ \varepsilon_{EU} + \lambda \left( \sum_{(k,j)} \left| x_{(i,j)}^h(k,l) - x_{(i,j)}^q(k,l) \right| \cdot \Delta(k,l) - \sum M_{(i,j)}^C(k,l) \right) \right\},
$$

where $\lambda$ is the multiplier. According to Eqs. (3), (15) and (16), we get

$$
\sum_{(k,j)} \text{sgn}(d) \cdot \delta_{(i,j)}(k,l) = \sum_{(k,j)} x_{(i,j)}^h(k,l) - \sum_{(k,j)} x_{(i,j)}^q(k,l) = \bar{S} - S = d;
$$

that is,

$$
\sum_{(k,j)} \delta_{(i,j)}(k,l) = |d|.
$$

We simplify the indices of $x_{i,j}^h(k,l)$, $x_{i,j}^q(k,l)$, $\Delta(k,l)$, and $M_{(i,j)}(k,l)$ by changing them to $x_s^h$, $x_s^q$, $\Delta_s$, and $M_s^C$, respectively, where the index $s$ is restricted to non-zero DCT coefficients. Then, Eq. (18) can be rewritten as

$$
\min_\delta \left[ L(\delta, \lambda) = \sum_{s=1}^{N_o} \left( (x_s^h - x_s^q) \cdot \Delta_s \right)^2 + \lambda \cdot \left( \sum_{s=1}^{N_o} |x_s^h - x_s^q| \cdot \Delta_s - \sum_{s=1}^{N_o} M_s^C \right) \right].
$$

(21)

Note that $x_s^h - x_s^q = \text{sgn}(d) \cdot \delta_s$ is a function of $\delta_s$. To obtain the optimal solution, we differentiate Eq. (21) with respect to $\delta$ and $\lambda$ and set each result to zero, i.e., $\partial L(\delta, \lambda)/\partial \delta_j = 0$ ($j=1,2,\ldots, N_o - 1$) and $\partial L(\delta, \lambda)/\partial \lambda = 0$. By using Eq. (20), the optimal solution $\tilde{\delta}^* = \{\delta_1^*, \delta_2^*, \ldots, \delta_{N_o-1}^*, \lambda^*\}^T$ can be calculated as

$$
\tilde{\delta}^* = \Delta_A^{-1} Q_A,
$$

where

$$
\Delta_A = \begin{pmatrix}
2\Delta_1^2 + 2\Delta_{N_o}^2 & 2\Delta_2^2 + 2\Delta_{N_o}^2 & 2\Delta_3^2 + 2\Delta_{N_o}^2 & \cdots & 2\Delta_{N_o-1}^2 + 2\Delta_{N_o}^2 & \Delta_1 - \Delta_{N_o} \\
2\Delta_2^2 + 2\Delta_{N_o}^2 & 2\Delta_3^2 + 2\Delta_{N_o}^2 & 2\Delta_4^2 + 2\Delta_{N_o}^2 & \cdots & 2\Delta_{N_o-2}^2 + 2\Delta_{N_o}^2 & \Delta_2 - \Delta_{N_o} \\
2\Delta_3^2 + 2\Delta_{N_o}^2 & 2\Delta_4^2 + 2\Delta_{N_o}^2 & 2\Delta_5^2 + 2\Delta_{N_o}^2 & \cdots & 2\Delta_{N_o-3}^2 + 2\Delta_{N_o}^2 & \Delta_3 - \Delta_{N_o} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
2\Delta_{N_o-2}^2 + 2\Delta_{N_o}^2 & 2\Delta_{N_o-1}^2 + 2\Delta_{N_o}^2 & 2\Delta_{N_o-3}^2 + 2\Delta_{N_o}^2 & \cdots & 2\Delta_{N_o-4}^2 + 2\Delta_{N_o}^2 & \Delta_{N_o-2} - \Delta_{N_o} \\
\Delta_1 - \Delta_{N_o} & \Delta_2 - \Delta_{N_o} & \Delta_3 - \Delta_{N_o} & \cdots & \Delta_{N_o-2} - \Delta_{N_o} & 0
\end{pmatrix}_{N_o \times N_o},
$$

(23)

$$
Q_A = \begin{pmatrix}
2 \cdot |d| \Delta_{N_o}^2 \\
2 \cdot |d| \Delta_{N_o}^2 \\
\vdots \\
2 \cdot |d| \Delta_{N_o}^2 \\
\sum_{i=1}^{N_o-1} M_i^C - |d| \Delta_{N_o}^2
\end{pmatrix}.
$$

(24)

Then, $\delta_{N_o}^*$ can be calculated based on Eq. (20), i.e., $\delta_{N_o}^* = |d| - \sum_{s=1}^{N_o-1} \delta_s^*$. The details are given in Appendix A.

**Case 2:** Insufficient capacity
In the case of a larger $L$, a higher $d^*$ and a higher level of modification $\{\delta_s|s=1,...,N_o\}$ are required. As the embedding capacity of an EU is insufficient, we adopt a different strategy for finding an initial solution. According to Eq. (10), the greater the magnitude of the JND mask, the lower the sensitivity of the human eye will be. Hence, the modification of each chosen DCT QIs can be adjusted according to the magnitude of its corresponding JND mask. Based on this concept, the optimization problem in this case can be expressed as

$$\min_\delta \left\{ L(\delta) = \sum_{s=1}^{N_o} \left( |x_s^h - x_s^p| \cdot \Delta_s - M_s^C \right)^2 \right\}. \tag{25}$$

By using Eq. (20), $L(\delta)$ in Eq. (25) can be rewritten as

$$L(\delta) = \sum_{s=1}^{N_o-1} (\delta_s \cdot \Delta_s - M_s^C)^2 + \left( \Delta_{N_o} \left( |d| - \sum_{s=1}^{N_o-1} \delta_s \right) - M_{N_o}^C \right)^2. \tag{26}$$

To obtain the optimal solution, we differentiate Eq. (26) with respect to $\delta$, and set each result to zero, i.e.,

$$\frac{\partial L(\delta)}{\partial \delta_j} = 0, j=1,2,...,N_o-1.$$ By using Eq. (20), the optimal solution $\vec{\delta}^*$ can be calculated as

$$\vec{\delta}^* = \Delta^\dagger \mathbf{Q}_\Delta, \tag{27}$$

where

$$\Delta = \begin{pmatrix} \Delta_1 \Delta_{N_o} \Delta_{N_o}^2 \cdots \Delta_{N_o}^2 \\ \Delta_{N_o} \Delta_2 \cdots \Delta_{N_o}^2 \\ \cdots \\ \Delta_{N_o} \cdots \Delta_{N_o}^2 + \Delta_{N_o}^2 \end{pmatrix}, \tag{28}$$

$$\mathbf{Q}_\Delta = \begin{pmatrix} |d|\Delta_{N_o}^2 + M_{N_o}^C \Delta_1 - M_{N_o}^C \Delta_{N_o} \\ |d|\Delta_{N_o}^2 + M_{N_o}^C \Delta_2 - M_{N_o}^C \Delta_{N_o} \\ \cdots \\ |d|\Delta_{N_o}^2 + M_{N_o}^C \Delta_{N_o-1} - M_{N_o}^C \Delta_{N_o} \end{pmatrix}. \tag{29}$$
Then, \( \delta_{N_o}^* \) can be obtained by using Eq. (20), i.e.,
\[
\delta_{N_o}^* = |d| - \sum_{s=1}^{N_o-1} \delta_s^*.
\]
The details of the derivations are given in Appendix B. Note that the initial solutions found by Eqs. (22) and (27) should be further rounded to the nearest integers before they are applied to their corresponding DCT QI’s.

B. Neighboring solutions

In the conventional SA algorithm, the “only one neighbor” solution is obtained by randomly altering the current solution. For our problem, this randomness is limited due to the inherent constraint in Eq. (19). Unlike the approach in [5], our method creates multiple neighboring solutions, and the best one is taken as the next current solution.

Multiple neighboring solutions can be created according to various rules (e.g., distributed, concentrated, and random), subject to the constraint in Eq. (19). The “concentrated” mode follows the observation in [4] about minimizing the MSE to improve picture quality. This shows that modifications concentrated on non-zero QIs of smaller quantization steps are much more effective. In [28], it is also shown that the smaller the number of altered QIs, the higher the undetectability of the stego image will be. Hence, concentrating the altered non-zero QI’s in the lower frequency band (where smaller quantization steps may occur) is helpful in achieving the optimum and speeding up the convergence. The “distributed” mode is supported by the argument in [16], which asserts that the sequential JSteg approach [15] causes obvious differences in the statistical properties of the altered part and the innocent part of a stego-image. However, the random JSteg approach, which distributes modifications among non-zero DCT QIs, is more secure.

Here, two random modification vectors are selected as the neighboring solutions. Thus, according to the above strategies, the search for neighboring solutions for SA generates four variations of the current solution. For example, if the current solution is \( \delta = [3,2,1,0,0,0,0] \) (as in case when \( d=8 \)), then we have the following neighboring solutions: \( \delta_1 = [2,2,2,1,0,0,0,0] \) (more even), \( \delta_2 = [4,2,2,0,0,0,0,0] \) (more centralized), \( \delta_3 = [2,2,2,2,0,0,0,0] \) (random), and \( \delta_4 = [3,3,1,1,0,0,0,0] \) (random). Note that each neighboring solution
should still satisfy the requirement of $\sum_{s=1}^{N_s} \delta_s = |d|$ (i.e., Eq. (20)). Because our SA uses neighboring solutions derived by different strategies (i.e., distributed, concentrated, and random), it is a promising way of searching for the optimal solution.

After obtaining the four neighboring solutions, their respective costs, calculated based on Eq. (1), are compared with the current solution. The solution with the lowest cost is then used to update the current solution for the next iteration. For the given parameters $T_{init}$, $T_{final}$, and $\Delta T$, the SA search procedure can be summarized as follows:

(a) Compute the initial cost $E^{(0)}$ based on an initial solution $\delta^p$, $p=0$.

(b) Find four neighboring solutions of $\delta^p$ and calculate their respective costs $E_u^p$ ($u=1, 2, 3, 4$).

(c) Compare $\min \{ E_u^p \}$ with $E^{(p-1)}$. Set

$$E^{(p)} = \min \{ E_u^p \}, \quad \text{if } \min \{ E_u^p \} \leq E^{(p-1)},$$

$$E^{(p)} = \min \{ E_u^p \}, \quad \text{if } \min \{ E_u^p \} \geq E^{(p-1)} \text{ and } \gamma < \exp \left\{ \frac{\min \{ E_u^p \} - E^{(p-1)}}{k_B T_{current}} \right\},$$

$$E^{(p)} = E^{(p-1)}, \quad \text{otherwise},$$

where $\gamma$ is a random number with a uniform distribution in $[0, 1]$; $\exp \{ \} \}$ denotes the exponential function; $k_B$ represents the Boltzmann constant; and $T_{current}$ is the current temperature. The condition in Eq.(30-2) is the so-called Metropolis rule, which leads to a movement towards a higher cost, but an escape from a trap.

(d) Increase $p$ by 1 and compare it with $N_I$ (maximum number of iterations). If $p$ is smaller than $N_I$, repeat Steps (b)–(d); otherwise, stop the iterative procedure.

Note that the anti-steganalysis test can be performed each time a current solution is updated. If the solution space is huge (e.g., a large $L$, or, a high embedding capacity) and a large number of iterations is
required for each EU, the anti-steganalysis test can be performed after that part of the search converges, as illustrated in Fig. 1.

V. EXPERIMENT RESULTS

To evaluate the performance of our proposed framework, we selected 200 JPEG images, each is of 512×512 pixels, as subjects for data embedding. The default quality factor for each JPEG image was 75%. Let an EU be 64×64 pixels and the number of message elements $N_m$ equal 64. The test message is randomly generated, where each element satisfies $0 \leq m_n < L$. The parameters of the SA algorithm are as follows: initial temperature $T_{\text{init}} = 200$, final temperature $T_{\text{final}} = 0$, decrement of temperature $\Delta T = 1$, and the number of iterations performed at each temperature is 100. The weights in the cost function are: $w_1=1$, $w_2=1$, and $w_3=10$ (meaning that anti-steganalysis enhancement is heavily emphasized). Actually, as mentioned in Section II, these weights can be set according to the user’s preferences. We conducted several experiments to evaluate the performances of the proposed algorithm in terms of imperceptibility, JPEG file-size variation, and anti-steganalysis pass-rate (defined as the ratio of the number of stego images not detected by the steganalyzer in [3] to the number of total test images). In addition to the traditional PSNR, NQM (noise quality metric) proposed in [12] was adopted to measure the picture quality based on an HVS model.

A. Performance of Imperceptibility

We first analyzed the impact of $L$ on the visual quality of images by varying it from 1024 to 8192, i.e., a capacity of 10 bits to 13 bits for each EU. The PSNRs of the modified JPEG images (with respect to their non-modified versions) were all higher than 41.98 dB, 38.89dB, 32.83dB and 27.94dB when $L$ was set to 1024, 2048, 4096, and 8192, respectively. These results demonstrate that the higher the embedding capacity (i.e., higher $L$), the lower the visual quality obtained. Figure 2
shows three popular cover images (Boat, Bridge, and Barbara) and the corresponding stego versions ($L=1024$) obtained by the SA algorithm. As we can see from Fig. 2, it is difficult to distinguish stego images from their cover versions with the human eye.

**B. Other performances of the proposed scheme**

Here we analyze the performance of SA. Figure 3 plots two curves of the cost function $E$ (i.e., Eq. (1)) for two separate EUs under the condition: $w_1=1$, $w_2=1$, (a) $w_3=1$ or (b) $w_3=10$, $T_{\text{init}} = 100$, $T_{\text{final}} = 0$, and $N_I = 20000$ (maximum number of iterations). It is shown that both curves decrease gradually as the iterations proceed. A jump-up and jump-down of cost function $E$ occur in Fig. 3(b). The former is due to the Metropolis rule (a very characteristic of the SA algorithm) in finding the global minimum. The latter comes from the goodness of the randomly selected neighboring solutions. The results demonstrate that a solution of low cost can be derived after a number of iterations based on the proposed SA-based scheme.

Fig. 2 (a) Three cover images and (b) the corresponding stego images ($L=1024$).
Fig. 3 The curves of cost function $E$ vs. iteration number with (a) $w_3=1$ and (b) $w_3=10$.

To evaluate the effect of the initial solution in SA, we compare the PSNRs of the stego results obtained from the proposed solution (Section IV.A) and the random initial solutions under the same embedding capacity ($L=1024$). The experiment results are listed in Table I. In the table, “No iterations” means the stego results based on the initial solutions only. The results show that, compared to random initial solutions with no iterations, the improvements in PSNR and NQM achieved by our proposed algorithm are 13.46 dB and 17.34 dB, respectively. Hence, our algorithm provides a good starting point for reducing the number of subsequent iterations or generating a higher peak.

The number of SA iterations is also crucial to converging to a good solution. As we can see from Table I, the lower the decrement of temperature $\Delta T$ (implying a larger number of iterations), the higher the improvement will be on PSNR or NQM for both cases of random and proposed initial solutions. These results manifest that a better solution can still be achieved after a large number of iterations, even if a random initial solution is used. For the case of intensive iterations ($\Delta T = 1$), our algorithm achieves an improvement of 8.64 dB (in PSNR) and 11.71 dB (in NQM) over the compared random method. On the other hand, an improvement of 7.37dB (PSNR), 7.39dB (NQM), and 25.5% (anti-steganalysis) can be also achieved for the
random initialization method by SA. These results manifest that the proposed SA-based framework effectively advances the performance of a steganographic scheme.

Table I Average PSNR, NQM, and anti-steganalysis pass-rate performances of the random and the proposed initial solutions for the complete test image set ($L=1024$).

<table>
<thead>
<tr>
<th></th>
<th>PSNR (dB)</th>
<th>NQM (dB) [12]</th>
<th>Anti-steganalysis pass-rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Proposed initial</td>
<td>Random initial</td>
<td>Proposed initial</td>
</tr>
<tr>
<td>No iterations</td>
<td>39.61</td>
<td>26.15</td>
<td>39.20</td>
</tr>
<tr>
<td>Iterations with $\Delta T = 5$</td>
<td>37.41</td>
<td>27.87</td>
<td>36.45</td>
</tr>
<tr>
<td>Iterations with $\Delta T = 1$</td>
<td>41.98</td>
<td>33.52</td>
<td>40.96</td>
</tr>
</tbody>
</table>

Readers might find it strange that the average PSNR and NQM in case of $\Delta T = 5$ are lower than those of the proposed initial solution. The reason is that we only consider imperceptibility (in terms of MSE and HVS indices) when we derive the initial solution, but other highly weighted factors, such as the anti-steganalysis property, dominate the search directions during the SA optimization process. As shown in Table I, the reductions in PSNR and NQM are traded for an increase in the anti-steganalysis pass-rate. This situation does not arise with the random initialization method because it lacks optimization in finding the initial solution.

With regard to the anti-steganalysis property, as shown in Table I, the pass-rate increases because SA performs more iterations, no matter what kind of initial solution is utilized. After a large number of iterations ($\Delta T = 1$), the improvements derived by our proposed method and the random initialization method is 24% and 25.5% respectively. Statistically, our average pass-rate, after the same number of iterations, is 12.5% higher than that of the random initialization method. In terms of picture quality and anti-steganalysis, a good initial solution is indeed helpful for finding a true optimum or reducing the number of iterations.

C. Performances under different weights $w_3$

As mentioned in Section II, weights in Eq. (1) can be adjusted according to user’s preferences. We change the weightings in Eq. (1) to $w_1=1$, $w_2=1$, and $w_3=1$ for the performance evaluation. It is expected that imperceptibility will regain its position. In the experiments, with $\Delta T = 5$, the PSNR, NQM, and pass-rate
were 39.73 dB, 38.06 dB, and 85.5% (not shown in tables), respectively, by using our proposed initial solution. This demonstrates that it is possible to tune the system performance (quality or anti-steganalysis) according to users’ preferences or requirements.

Here we analyze a special case $w_3=0$, i.e., blind anti-steganalysis. The result is given in Table II, where $w_1=1$, $w_2=1$, $w_3=0$, and $\Delta T = 1$. As we can see from Table II, the PSNR and NQM of the developed scheme with the proposed initial solutions are 42.35 dB and 41.62 dB, respectively, which are higher than those listed in Table I. On the other hand, the anti-steganalysis pass-rate is reduced, but still maintained above 80%, even if the information about the possible steganalyzers is not given. A similar behavior occurs for the random initial solutions with $w_3=0$. These results show that it is difficult to make large improvements on anti-steganalysis with a full blindness to steganalyzers, but the SA-based optimization mechanism indeed works to raise the imperceptibility and anti-steganalysis simultaneously if the possible steganalytic schemes are known as a priori.

<table>
<thead>
<tr>
<th>Table II</th>
<th>Average PSNR, NQM, and anti-steganalysis pass-rate performances of the random and the proposed initial solutions for the complete test image set ($L=1024$) under blind anti-steganalysis ($w_3=0$).</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PSNR (dB)</td>
</tr>
<tr>
<td></td>
<td>Proposed initial</td>
</tr>
<tr>
<td>Iterations with $\Delta T = 1$</td>
<td>42.35</td>
</tr>
</tbody>
</table>

D. Comparison with existing method [5]

The steganographic method described in [5] also adopts the sum of DCT QIs (i.e., Eq. (3)) to hide message bits, but it is an open-loop method. Here, we compare the PSNR, file-size variation, and NQM of the proposed method and the method in [5]. To be fair, we set the embedding capacity to be the same for both methods. As shown in Table III, the proposed scheme improves the PSNRs by an average of 6.68 dB. In addition, the NQMs of the proposed scheme are of 2.32 dB higher than those in [5]. This is because the
impact of the quantization step size and the HVS effect on the visual quality are considered by the performance indices $f_1$ and $f_2$ during closed-loop operation.

**Table III** Comparison of the PSNR, file-size variation, NQM, and anti-steganalysis pass-rate performance of the proposed scheme and the method in [5].

<table>
<thead>
<tr>
<th>$L$</th>
<th>PSNR (dB)</th>
<th>Variation in file-size</th>
<th>NQM (dB) [12]</th>
<th>Anti-steganalysis Pass rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1024</td>
<td>41.98</td>
<td>36.77</td>
<td>0.30%</td>
<td>6.11%</td>
</tr>
<tr>
<td>2048</td>
<td>38.89</td>
<td>30.82</td>
<td>0.89%</td>
<td>18.62%</td>
</tr>
<tr>
<td>4096</td>
<td>32.83</td>
<td>25.37</td>
<td>3.61%</td>
<td>43.88%</td>
</tr>
<tr>
<td>8192</td>
<td>27.94</td>
<td>22.15</td>
<td>9.02%</td>
<td>83.72%</td>
</tr>
</tbody>
</table>

The JPEG file size would probably be modified due to hidden message data. Variations in the size of JPEG files are 3.46% and 38.08% on average for our method and [5], respectively. The superior performance of our method is due to the statistical feature $f_3$ evaluated in the loop, which ensures that the distribution of DCT QIs remains virtually unchanged. Therefore, in terms of both image quality and file-size variation, our scheme outperforms [5] by including the SA-based optimization in a closed-loop architecture.

Next, we evaluate the anti-steganalysis performance of both methods. In this experiment, the embedding capacities of the stego images generated by the two schemes are kept the same in order to ensure fair comparison. From Table III, we observe that the pass-rate of the proposed scheme is substantially higher than that in [5] as $L$ is 4096 (100% vs. 9%). This shows that our scheme has substantially enhanced the anti-steganalysis capability of the steganographic scheme proposed in [5], as the embedding capacity becomes higher than $L=4096$. The main reason is that the performance index $f_3$ for measuring the variation between the statistical properties of the cover and the stego images is adopted in the loop; hence it is kept small by using SA-based optimization. However, the pass-rate is only 90.5% when $L$ is equal to 8192. This is due to a heavy embedding rate, which cannot achieve convergence at the targeted number of iterations. Adjusting the parameters, e.g., $T_{\text{init}}$, $T_{\text{final}}$, and $\Delta T$, to increase the number of iterations allowed in the SA search might be the only way to solve this problem.
VI. CONCLUSIONS AND REMARKS

We have proposed a SA-based framework to optimize the visual quality of stego images and enhance the anti-steganalysis property simultaneously. Note that the goal of this paper is to develop such an optimizing framework, but not intently for any particular stegnography or steganalysis algorithm. The experiment results show that the proposed closed-loop system has improved the picture quality by 6.68 dB with respect to the base algorithm in [5] and enhanced the anti-steganalysis property (100% pass-rate for $L \leq 4096$) after data embedding. In principle, this framework can be applied to other steganographic schemes and other steganalytic systems. Moreover, extra constraints or indices other than picture quality, HVS, and anti-steganalysis can be added to the cost function for multi-objective optimization.

Finally, we draw two remarks below.

(1) Experiments show that blind anti-steganalysis is indeed difficult and even the SA algorithm can advance the pass-rate by only 2.5%~8.5%. However, with a matched hypothesis about steganalysis (certainly, $w_3 \neq 0$), the advantage of SA on anti-steganalysis can be up to 24%. Based on the proposed framework, users can further enrich the set of feature $f_3$ so that resistance to a wider range of steganalysis algorithms can be achieved.

(2) Our framework is not universal to all types of steganographical algorithms. For steganographical algorithms that embed one or more bits for each pixel (instead of a region), our framework may not work. The premise of our framework is that there should be many manners for embedding a single bit (or a plurality of bits) and the manner is to be optimized.

APPENDIX A

Here, we derive the optimal solution to Eq. (21) in detail. Substituting Eq. (14) into Eq. (21), $L(\delta, \lambda)$ can be rewritten as
\[
L(\delta, \lambda) = \sum_{s=1}^{N_s} \left( (x_s^b - x_s^q) \cdot \Delta_s \right)^2 + \lambda \left( \sum_{s=1}^{N_s} |\delta_s| \cdot \Delta_s - \sum_{s=1}^{N_s} M_s^C \right)
= \sum_{s=1}^{N_s} \left( \text{sgn}(d) \cdot \delta_s \cdot \Delta_s \right)^2 + \lambda \left( \sum_{s=1}^{N_s} |\delta_s| \cdot \Delta_s - \sum_{s=1}^{N_s} M_s^C \right)
= \sum_{s=1}^{N_s} (\delta_s \cdot \Delta_s)^2 + \lambda \left( \sum_{s=1}^{N_s} \delta_s \cdot \Delta_s - \sum_{s=1}^{N_s} M_s^C \right). 
\] (A.1)

Combining Eq. (20) and Eq. (A.1), we obtain

\[
L(\delta, \lambda) = \sum_{s=1}^{N_s-1} (\delta_s \cdot \Delta_s)^2 + \left( |d| - \sum_{s=1}^{N_s-1} \delta_s \right) \Delta_{N_s}^2 + \lambda \left( \sum_{s=1}^{N_s-1} \delta_s \cdot \Delta_s - \sum_{s=1}^{N_s} M_s^C \right)
+ \lambda \left( |d| - \sum_{s=1}^{N_s-1} \delta_s \right) \Delta_{N_s}^2,
\] (A.2)

Differentiating Eq. (A.2) with respect to \(\delta\) (i.e., \(\partial L(\delta, \lambda) / \partial \delta\)) yields

\[
\frac{\partial L(\delta, \lambda)}{\partial \delta_k} = 2\delta_k \Delta_k^2 - 2 \left( |d| - \sum_{s=1}^{N_s-1} \delta_s \right) \Delta_{N_s}^2 + \lambda \left( \Delta_k - \Delta_{N_s} \right)
= 2\delta_k \Delta_k^2 + 2\Delta_{N_s}^2 \sum_{s=1}^{N_s-1} \delta_s - 2|d| \Delta_{N_s}^2 + \lambda \left( \Delta_k - \Delta_{N_s} \right),
\] (A.3)

where \(k = 1, 2, \ldots, N_s - 1\). On the other hand, differentiating Eq. (A.2) with respect to \(\lambda\) (i.e., \(\partial L(\delta, \lambda) / \partial \lambda\)) yields

\[
\frac{\partial L(\delta, \lambda)}{\partial \lambda} = \sum_{s=1}^{N_s-1} (\delta_s \Delta_s) + \Delta_{N_s} \left( |d| - \sum_{s=1}^{N_s-1} \delta_s \right) - \sum_{s=1}^{N_s} M_s^C.
\] (A.4)

To obtain the optimal solution, we set Eqs. (A.3) and (A.4) to zero, i.e., \(\partial L(\delta, \lambda) / \partial \delta = 0\) and \(\partial L(\delta, \lambda) / \partial \lambda = 0\), and express them in the following matrix form:

\[
\begin{pmatrix}
2\Delta_1^2 + 2\Delta_{N_s}^2 & 2\Delta_2^2 + 2\Delta_{N_s}^2 & \cdots & 2\Delta_{N_s-1}^2 + 2\Delta_{N_s}^2 & \Delta_1 - \Delta_{N_s} \\
2\Delta_{N_s}^2 & 2\Delta_2^2 + 2\Delta_{N_s}^2 & \cdots & 2\Delta_{N_s-1}^2 + 2\Delta_{N_s}^2 & \Delta_2 - \Delta_{N_s} \\
2\Delta_2^2 & 2\Delta_3^2 + 2\Delta_{N_s}^2 & \cdots & 2\Delta_{N_s-1}^2 + 2\Delta_{N_s}^2 & \Delta_3 - \Delta_{N_s} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
2\Delta_{N_s}^2 & 2\Delta_{N_s}^2 & \cdots & 2\Delta_{N_s-1}^2 + 2\Delta_{N_s}^2 & \Delta_{N_s-1} - \Delta_{N_s} \\
\Delta_1 - \Delta_{N_s} & \Delta_2 - \Delta_{N_s} & \cdots & \Delta_{N_s-1} - \Delta_{N_s} & 0
\end{pmatrix}
\begin{pmatrix}
\delta_1 \\
\delta_2 \\
\delta_3 \\
\vdots \\
\delta_{N_s-1} \\
\lambda
\end{pmatrix}
= \begin{pmatrix}
2|d| \Delta_{N_s}^2 \\
2|d| \Delta_{N_s}^2 \\
2|d| \Delta_{N_s}^2 \\
\vdots \\
2|d| \Delta_{N_s}^2 \\
\sum_{s=1}^{N_s} M_s^C - |d| \Delta_{N_s}
\end{pmatrix}.
\] (A.5)
By defining

\[
\Delta_A = \begin{pmatrix}
2\Delta_1^2 + 2\Delta_{N_o}^2 & 2\Delta_{N_o}^2 & 2\Delta_{N_o}^2 & \cdots & 2\Delta_{N_o}^2 & \Delta_1 - \Delta_{N_o} \\
2\Delta_{N_o}^2 & 2\Delta_2^2 + 2\Delta_{N_o}^2 & 2\Delta_{N_o}^2 & \cdots & 2\Delta_{N_o}^2 & \Delta_2 - \Delta_{N_o} \\
2\Delta_{N_o}^2 & 2\Delta_{N_o}^2 & 2\Delta_3^2 + 2\Delta_{N_o}^2 & \cdots & 2\Delta_{N_o}^2 & \Delta_3 - \Delta_{N_o} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
2\Delta_{N_o}^2 & 2\Delta_{N_o}^2 & 2\Delta_{N_o}^2 & \cdots & 2\Delta_{N_o}^2 & \Delta_{N_o-1} - \Delta_{N_o} \\
\Delta_1 - \Delta_{N_o} & \Delta_2 - \Delta_{N_o} & \Delta_3 - \Delta_{N_o} & \cdots & \Delta_{N_o-1} - \Delta_{N_o} & 0
\end{pmatrix}_{N_o \times N_o},
\]  

(A.6)

\[
\tilde{\delta} = \begin{bmatrix} \delta_1 & \delta_2 & \cdots & \delta_{N_o-1} & \lambda \end{bmatrix}^T,
\]

and

\[
Q_A = \begin{pmatrix}
2 \cdot |k| \Delta_{N_o}^2 \\
2 \cdot |k| \Delta_{N_o}^2 \\
2 \cdot |k| \Delta_{N_o}^2 \\
\vdots \\
2 \cdot |k| \Delta_{N_o}^2 \\
\left( \sum_{s=1}^{N_o-1} M_s \right) - |d| \Delta_{N_o}
\end{pmatrix}_{N_o \times 1}.
\]

(A.7)

Eq. (A.5) can be expressed as

\[
\Delta_A \tilde{\delta} = Q_A,
\]

(A.8)

Then we can compute the optimal solution as

\[
\tilde{\delta}^* = \Delta_A^{-1} Q_A.
\]

(A.9)

**APPENDIX B**

Here, we derive the optimal solution to Eq. (25) in detail. For clarity, we rewrite Eq. (25) as follows:

\[
L(\delta) = \sum_{s=1}^{N_o} \left( |x_h^s - x_q^s| \cdot \Delta_s - M_s^c \right)^2 \\
= \sum_{s=1}^{N_o} \left( \text{sgn}(d) \delta_s \cdot \Delta_s - M_s^c \right)^2 \\
= \sum_{s=1}^{N_o} \left( \delta_s \cdot \Delta_s - M_s^c \right)^2.
\]

(B.1)

Combining Eqs. (20) and (B.1), we obtain
\[ L(\delta) = \sum_{s=1}^{N_o-1} \left( \delta_s \cdot \Delta_s - M_s^C \right)^2 + \left( \delta_{N_o} \cdot \Delta_{N_o} - M_{N_o}^C \right)^2 \]  
\[ = \sum_{s=1}^{N_o-1} \left( \delta_s \cdot \Delta_s - M_s^C \right)^2 + \left( \Delta_{N_o} \left( |d| - \sum_{s=1}^{N_o-1} \delta_s \right) - M_{N_o}^C \right)^2. \]  

(B.2)

Differentiating Eq. (B.2) with respect to \( \delta \) (i.e., \( \partial L(\delta) / \partial \delta \)) yields

\[ \frac{\partial}{\partial \delta_k} L(\delta) = 2(\delta_k \Delta_k - M_k^C) \cdot \Delta_k + 2 \left( |d| - \sum_{s=1}^{N_o-1} \delta_s \right) \left( - \Delta_{N_o} \right), \quad k=1,2,\ldots,N_o-1. \]  

(B.3)

To obtain the optimal solution, we set Eq. (B.3) to zero, i.e., \( \frac{\partial}{\partial \delta_k} L(\delta) = 0 \), and get

\[ \delta_k \Delta_k^2 + \Delta_{N_o}^2 \sum_{s=1}^{N_o-1} \delta_s = |d| \Delta_{N_o}^2 + M_k^C \Delta_k - M_{N_o}^C \Delta_{N_o}, \quad k=1,2,\ldots,N_o-1. \]  

(B.4)

Eq. (B.4) can be expressed in the following matrix form:

\[ \begin{pmatrix}
\Delta_1^2 + \Delta_{N_o}^2 & \Delta_{N_o}^2 & \cdots & \Delta_{N_o}^2 \\
\Delta_{N_o}^2 & \Delta_2^2 + \Delta_{N_o}^2 & \cdots & \Delta_{N_o}^2 \\
\vdots & \vdots & \ddots & \vdots \\
\Delta_{N_o}^2 & \Delta_{N_o}^2 & \cdots & \Delta_{N_o}^2 + \Delta_{N_o}^2 \\
\end{pmatrix}
\begin{pmatrix}
\delta_1 \\
\delta_2 \\
\vdots \\
\delta_{N_o-1} \\
\end{pmatrix}
= \begin{pmatrix}
|d| \Delta_{N_o}^2 + M_1^C \Delta_1 - M_{N_o}^C \Delta_{N_o} \\
|d| \Delta_{N_o}^2 + M_2^C \Delta_2 - M_{N_o}^C \Delta_{N_o} \\
\vdots \\
|d| \Delta_{N_o}^2 + M_{N_o-1}^C \Delta_{N_o-1} - M_{N_o}^C \Delta_{N_o} \\
\end{pmatrix}. \]  

(B.5)

By defining

\[ \Delta_\Lambda = \begin{pmatrix}
\Delta_1^2 + \Delta_{N_o}^2 & \Delta_{N_o}^2 & \cdots & \Delta_{N_o}^2 \\
\Delta_{N_o}^2 & \Delta_2^2 + \Delta_{N_o}^2 & \cdots & \Delta_{N_o}^2 \\
\vdots & \vdots & \ddots & \vdots \\
\Delta_{N_o}^2 & \Delta_{N_o}^2 & \cdots & \Delta_{N_o}^2 + \Delta_{N_o}^2 \\
\end{pmatrix}, \]  

(B.6),

\[ \delta = [\delta_1 \quad \delta_2 \quad \ldots \quad \delta_{N_o-1}]^T, \]

and

\[ \mathbf{Q}_\Lambda = \begin{pmatrix}
|d| \Delta_{N_o}^2 + M_1^C \Delta_1 - M_{N_o}^C \Delta_{N_o} \\
|d| \Delta_{N_o}^2 + M_2^C \Delta_2 - M_{N_o}^C \Delta_{N_o} \\
\vdots \\
|d| \Delta_{N_o}^2 + M_{N_o-1}^C \Delta_{N_o-1} - M_{N_o}^C \Delta_{N_o} \\
\end{pmatrix}. \]  

(B.7)

Eq. (B.5) can be expressed as
\[ \Delta_{A}\tilde{\delta} = Q_{A}, \]  
(B.8)

Then, the following optimal solution of \( \tilde{\delta} \) can be obtained:

\[ \tilde{\delta}^* = \Delta_{A}^{-1}Q_{A}, \]  
(B.9)

After calculating \( \tilde{\delta}_{k}^* (k=1,2,\ldots,N_{o} - 1) \), \( \tilde{\delta}^*_{N_{o}} \) can be computed via the relationship in Eq. (20).

REFERENCES