QUALITY-OPTIMIZED IMAGE STEGANOGRAPHY SUBJECT TO ANTI-STEGANALYSIS CONSTRAINT

Guo-Shiang Lin[1], *Yi-Ting Chang*[2], *Wen-Nung Lie*[2]

[1]Dept. of Computer Science and Information Engineering, Da-Yeh University, Chang-Hua, Taiwan, R.O.C.
khlin@mail.dyu.edu.tw
[2]Dept. of Electrical Engineering, National Chung Cheng University, Chia-Yi, Taiwan, R.O.C.
wnlie@ee.ccu.edu.tw

ABSTRACT

In this paper, we propose an architecture that combines a quantization-based steganogrphic scheme with a steganalysis system, operated in a closed-loop manner with a cost function for minimization, to enhance the anti-steganalysis capability and image quality after data embedding. In this architecture, a controller based on the SA (simulated annealing) technique is adopted to guarantee fast and accurate convergence. Our proposed system is applied to the data embedding of JPEG-compressed images. Compared with the original embedding algorithm in [5], a better image quality (by an average improvement of 6.48 dB) can be achieved and simultaneously the anti-steganalysis capability is enhanced significantly.

Index Terms— Steganography, steganalysis, HVS

1. INTRODUCTION

For steganographic schemes, imperceptibility is clearly the most important requirement [1]. That is, the modifications between the cover media and its stego version should be slight and transparent to the human eyes. However, today, this slight modification might be discoverable by using an adequate mechanism with the aid of computer, e.g., steganalysis [2][3]. Most steganalytic schemes analyze statistical properties to distinguish stego media from the original ones. Due to the above reason, some few researchers [11][12] pay more attention to this additional requirement of statistical undetectability to develop a mechanism against steganalysis. This technique, still in its infant, was called "anti-steganalysis."

Wu et al. [11] proposed a scheme in which a genetic algorithm (GA) is used to iteratively modify pixel-domain graylevels and make the difference of statistical properties between the cover image and its stego version small. The process will terminate until the steganalytic scheme in the loop fails and the bit error rate (BER) after data extraction is minimized. However, their steganographic scheme can not guarantee a BER of zero and the anti-steganalysis is based on each 8×8 pixel block, but not a full frame, implying a possible failure for attacks that consider statistic features of more than one block. On the other hand, Lie et. al. [4] developed an embedding algorithm for JPEG images to optimize the modified picture quality at a given robustness. A criterion of minimizing least-square-errors was adopted to calculate the modified amount to each transform coefficient, while maintaining the BER to zero in case of no attacks (e.g., common image processing). However, their work did not consider the steganalysis attack. The above drawbacks motivate us to develop an embedding architecture to solve the problems of anti-steganalysis, optimized picture quality, and BER, simultaneously.

Here in this paper, we propose an SA (Simulated Annealing)-based algorithm to augment a known [5] steganographic scheme for modifying transform coefficients so that some performance indices (e.g., MSE, HVS deviation, and differences of statistical features) are optimized subject to certain specified constraints (e.g., BER and anti-steganalysis).

2. THE SYSTEM ARCHITECTURE

Several factors are often considered in designing a stegnographic scheme [1]: (1) imperceptibility, (2) statistical undetectability (i.e., anti-steganalysis), (3) embedding capacity, and (4) bit error rate (BER) after data extraction. However, some of them might contradict to one another. For example, increasing the embedding capacity might decrease the imperceptibility and increase the statistical detectability. It is thus difficult to tradeoff them based on an open-loop architecture. Here, we propose a closed-loop architecture to achieve the above purposes simultaneously.

Figure 1 shows the architecture of our proposed system. Essentially, four kinds of performances are evaluated and used to steer the modification of transform coefficients: MSE (mean square error) f_1, HVS deviation f_2, difference of statistical features (for steganalysis) f_3, and BER f_4. To successfully pass the steganalysis attack for the resulting images, a steganalytic subsystem [3] from our prior work, which analyzes the gradient energy (in the spatial domain) and the Laplacian parameter (in the DCT transform domain) as the statistical features, is placed in the evaluation loop. The index f_3 measures the sum of absolute differences of

these two statistical features between a host image and its stego version. Then we define a cost function E as follows.

$$E = w_1 \times f_1 + w_2 \times f_2 + w_3 \times f_3 + w_4 \times f_4 \qquad (1)$$

where w_1, w_2, w_3, and w_4 are predefined weightings.

It is expected that the statistical ranges of these four features are different. In order to balance the effect of each feature in the cost function, they should be normalized. The normalization is defined as follows.

$$\hat{f}_i = \left(f_i - f_{i,\min}\right) / \left(f_{i,\max} - f_{i,\min}\right) \qquad (2)$$

where f_i and $\hat{f}_i$ denote the original feature and its normalized value, respectively, i =1,2,3,4, $f_{i,\min}$ and $f_{i,\max}$ represent the minimum and maximum values of the i-th feature, respectively.

To achieve convergence efficiently, optimization search is conducted by the SA (Simulated Annealing) mechanism [6], with good initial solutions estimated by some analytic techniques.

Since most digital images are stored and transmitted in JPEG format, here we pay more attention to combining the DCT-domain embedding schemes in [5] with our closed-loop architecture. Theoretically, this architecture can be extended to other steganographic schemes.

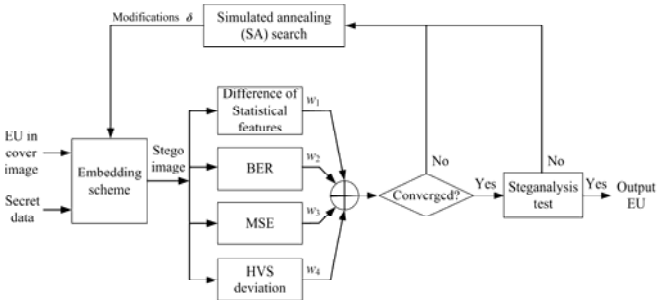

Fig. 1 Proposed closed-loop architecture for SA-based data embedding with enhanced anti-steganalysis.

3. THE STEGANOGRAPHIC SCHEME

First of all, an integer L should be determined before data embedding. Actually, L determines the embedding capacity of each embedding unit (EU), which is $\log_2 L$ bits. It is assumed that L is known to both the transmitter and the receiver.

Suppose the secret message $\boldsymbol{M}$ consists of N_m elements and is represented as $\boldsymbol{M} = \{m_n \mid n=1,2,..., N_m\}$. Message element m_n is hidden into the n-th EU and $0 \le m_n < L$. The embedding process is repeated N_m times to hide $\boldsymbol{M}$ into the whole image. To deal with JPEG images, the quantization index (QI) of DCT coefficients is used for data hiding. To have a high embedding efficiency and less variation in the resulting JPEG file size, only non-zero QIs are chosen for modification. Then the embedding algorithm is described as follows [5].

H1. Get the DCT QIs of the given JPEG image after VLD (Variable Length Decoding) process.

H2. Divide all DCT blocks into N_m EUs. Each EU contains several 8×8 DCT blocks.

H3. Calculate the sum S of the DCT QI in the n-th EU as:

$$S = \sum_{\substack{(k,l)\in\Omega' \\ (i,j)\in \mathrm{EU}}} x_{(i,j)}(k,l) \qquad (3)$$

$$r = S \bmod L,\ r \ge 0 \qquad (4)$$

where $\boldsymbol{X}_{i,j} = \{x_{(i,j)}(k,l) \mid 0 \le k,l \le 7\}$ is the DCT block representation of the (i,j)-th block, (k, l) is the position index of QI in Ω' (the set of non-zero QIs), and mod represents the modulo operator. Obviously, we have the remainder $0 \le r \le L-1$.

H4. Compute the modification amount d for data embedding as

$$d_1 = |m_n - r|,\ d_2 = L - d_1, \qquad (5)$$

and

$$d = \min(d_1, d_2). \qquad (6)$$

H5. Modify the DCT QI's in the n-th EU such that the following rules are satisfied:

$$\begin{aligned}
\tilde{S} &= S - d, && \text{if } r > m_n \text{ and } d_1 < d_2 \\
\tilde{S} &= S + d, && \text{if } r > m_n \text{ and } d_1 \ge d_2 \\
\tilde{S} &= S + d, && \text{if } r < m_n \text{ and } d_1 < d_2 \\
\tilde{S} &= S - d, && \text{if } r < m_n \text{ and } d_1 \ge d_2 \\
\tilde{S} &= S, && \text{if } r = m_n
\end{aligned} \qquad (7)$$

where $\tilde{S}$ denotes the modified version of S.

H6. Repeat Steps H1-H5 for all EUs to create a stego image.

To extract the hidden message data, the steps are:

E1. Decode the received JPEG image to obtain the DCT QIs and divide all DCT blocks into N_m EUs.

E2. Calculate $\hat{S}$ for each EU according to Eq.(3). Compute $\hat{r} = \hat{S} \bmod L$ (L is assumed known). Then $\hat{r}$ represents the hidden message m_n.

E3. Repeat Steps E1-E2 for all EUs to get all message data.

A problem not mentioned is how to alter DCT QIs in Step (H5) such that $|\tilde{S} - S| = d$ is satisfied. Several solutions are possible. For example, d is evenly distributed between all non-zero QIs. However, it is our intent to advance this steganographic scheme by properly distributing d among non-zero coefficients so that the cost function E in Eq.(1) is optimized and the steganalytic system is broken after embedding. Expectedly, the computational complexity of finding solutions by using the exhaustive search is extremely high. Here, a more efficient manner based on the SA algorithm is adopted instead. The detailed SA algorithm will not be addressed here due to space limitation. Readers can refer to related textbook or articles [6]. In short, three

important elements should be contained in SA: initial solution, neighboring solutions, and cost function.

4. SA-BASED OPTIMIZATION

Clearly, the cost function defined in Eq.(1) fits this purpose. Note that due to the nature of the embedding algorithm we adopt, BER will be always zero without external processing attacks, such as filtering, compression, etc. This is different from the case of [11], where BER cannot be guaranteed even after GA optimization. Hence, we set w_4=0 in current implementation.

As mentioned in [6], a good initial solution results in a faster convergence. Here, an analytic solution based on the minimization of MSE (f_1), subject to the HVS model (f_2), is estimated and used as the initial search.

First, $\boldsymbol{\delta} = \{\delta_{(i,j)}(k,l) \mid \delta_{(i,j)}(k,l) \geq 0, (k,l) \in \Omega'_{(i,j)}\}$ denotes the vector of modifications on QIs for each EU, where $\Omega'_{(i,j)}$ is the set of nonzero DCT QIs in the (*i,j*)-th 8×8 block. Hence, we have [4]:

$$x^h_{(i,j)}(k,l) = x^o_{(i,j)}(k,l) + \mathrm{sgn}(d) \cdot \delta_{(i,j)}(k,l), \qquad (8)$$

and

$$\sum_{\substack{(k,l)\in\Omega' \\ (i,j)\in \mathrm{EU}}} x^h_{(i,j)}(k,l) = \widetilde{S} \qquad (9)$$

where $\mathrm{sgn}(\cdot)$ is the sign function. To achieve imperceptibility, the Watson's perceptual model [8][9] is used here. Thus, the problem, minimizing MSE subject to HVS model, can be expressed as

$$\begin{aligned} &\min\ \varepsilon_{\mathrm{EU}} = \sum_{\substack{(k,l)\in\Omega' \\ (i,j)\in \mathrm{EU}}} (x^h_{(i,j)}(k,l) - x^o_{(i,j)}(k,l))^2 \ \text{ subject to} \\ &\sum_{\substack{(k,l)\in\Omega' \\ (i,j)\in \mathrm{EU}}} \left| x^h_{(i,j)}(k,l) - x^o_{(i,j)}(k,l) \right| \cdot \Delta(k,l) < \sum_{\substack{(k,l)\in\Omega' \\ (i,j)\in \mathrm{EU}}} M_{(i,j)}(k,l) \end{aligned}, \qquad (10)$$

where $\Delta(k,l)$ is the quantization step for the (*k,l*)-th coefficient, $M_{(i,j)}(k,l)$ is the mask value subject to HVS model [8][9]. Note that we have integrated the HVS constraints (i.e., $M_{(i,j)}(k,l)$) for individual coefficients into only one, due to the requirement of problem simplification. The constrained optimization problem stated in Eq. (10) can be converted into an unconstrained optimization problem by using the Lagrangian multiplier method. Then, Eq.(10) can be expressed as

$$\begin{aligned} &\min_{\boldsymbol{\delta}}\ L(\boldsymbol{\delta}, \lambda) \\ &= \min_{\boldsymbol{\delta}} \left\{ \varepsilon_{\mathrm{EU}}(\boldsymbol{\delta}) + \lambda \cdot \left(\sum_{i=1}^{N_o} \left| x^h_i - x^o_i \right| \cdot \Delta_i < \sum_{i=1}^{N_o} M_i \right) \right\} \end{aligned} \qquad (11)$$

where N_o denotes the number of nonzero DCT QIs in this EU. Without loss of clarity, we simplify the indices of $x^h_{(i,j)}(k,l)$, $x^o_{(i,j)}(k,l)$, and $M_{(i,j)}(k,l)$ and change them to be x^h_i, x^o_i, and M_i, respectively.

To obtain the optimal solution, we differentiate Eq.(11) with respect to $\boldsymbol{\delta}$ and λ, and set the result to zero, i.e., $\partial L(\boldsymbol{\delta}, \lambda) / \partial \boldsymbol{\delta} = 0$ and $\partial L(\boldsymbol{\delta}, \lambda) / \partial \lambda = 0$. Then the optimal solution $\widetilde{\boldsymbol{\delta}}^*$ can be calculated as follows:

$$\widetilde{\boldsymbol{\delta}}^* = \boldsymbol{A}^{-1}\boldsymbol{c}, \qquad (12)$$

where

$$\boldsymbol{A} = \begin{pmatrix} 2\Delta_1^2 + 2\Delta_{N_o}^2 & 2\Delta_{N_o}^2 & 2\Delta_{N_o}^2 & \cdots & 2\Delta_{N_o}^2 & \Delta_1 - \Delta_{N_o} \\ 2\Delta_{N_o}^2 & 2\Delta_2^2 + 2\Delta_{N_o}^2 & 2\Delta_{N_o}^2 & \cdots & 2\Delta_{N_o}^2 & \Delta_2 - \Delta_{N_o} \\ 2\Delta_{N_o}^2 & 2\Delta_{N_o}^2 & 2\Delta_3^2 + 2\Delta_{N_o}^2 & \cdots & 2\Delta_{N_o}^2 & \Delta_3 - \Delta_{N_o} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 2\Delta_{N_o}^2 & 2\Delta_{N_o}^2 & 2\Delta_{N_o}^2 & \cdots & 2\Delta_{N_o-1}^2\Delta_{N_o}^2 & \Delta_{N_o-1} - \Delta_{N_o} \\ \Delta_1 - \Delta_{N_o} & \Delta_2 - \Delta_{N_o} & \Delta_3 - \Delta_{N_o} & \cdots & \Delta_{N_o-1} - \Delta_{N_o} & 0 \end{pmatrix},$$

$$\widetilde{\boldsymbol{\delta}}^* = \begin{pmatrix} \delta_1^* \\ \delta_2^* \\ \delta_3^* \\ \vdots \\ \delta_{N_o-1}^* \\ \lambda^* \end{pmatrix} \text{ and } \boldsymbol{c} = \begin{pmatrix} 2 \mid d \mid \Delta_{N_o}^2 \\ 2 \mid d \mid \Delta_{N_o}^2 \\ 2 \mid d \mid \Delta_{N_o}^2 \\ \vdots \\ 2 \mid d \mid \Delta_{N_o}^2 \\ (\sum_{i=1}^{N_o-1} M_i) - \mid d \mid \Delta_{N_o} \end{pmatrix}$$

Then the solution can be utilized as the initial solution of SA. On searching neighboring solutions for SA, 4 variations of current solution, according to different strategies, are generated. As an example, if current solution is $\boldsymbol{\delta}$ =[3,2, 2,1,0,0,0,0] (i.e., *d*=8), then we can have neighboring solutions as $\boldsymbol{\delta}_1$=[2,2,2,1,1,0,0,0] (more even), $\boldsymbol{\delta}_2$=[4,2,2,0, 0,0,0,0] (more centralized), $\boldsymbol{\delta}_3$=[2,2,2,2,0,0,0,0] (random), and $\boldsymbol{\delta}_4$ =[3,3,1,1,0,0,0,0] (random). Note that each neighboring solution should still satisfy the requirement of $|\widetilde{S} - S| = d$. With neighboring solutions of different strategies, it is more possible for SA to search out the optimal one. Neighboring solutions, after evaluation, are then compared with the current solution and update it if necessary.

Note also that the steganalysis test can be done each time a current solution is updated or the search is converged, as illustrated in Fig.1. The requirement of anti-steganalysis will be released after a large times of iterations for each EU. This case often occurs at a large *L*, or a higher embedding capacity (see Table I).

5. EXPERIMENTAL RESULTS

Here 50 JPEG images, each of 512×512 pixels, were selected as subjects for data embedding. Let an EU is of 64×64 pixels. N_m equals 64 for each cover image.

We first analyze the impact of *L* on images' visual quality by varying it from 1024 to 8192, i.e., a capacity of 10 bits to 13 bits for each EU. The PSNRs measured between the input and the marked JPEG images are higher than 42.13 dB, 38.24dB, 31.94dB and 27dB when *L* is set to 1024, 2048, 4096, and 8192, respectively. The results show

that the higher the embedding capacity (i.e., higher L), the lower the visual quality.

Here we compare the PSNR, file size variance, and NQM (visual quality metric in [10]) of the proposed method with that in [5]. To be fair, the embedding rate is same to both. As shown in Table I, the proposed scheme improves the PSNRs by an average of 6.48 dB. As mentioned above, the JPEG file size would probably be modified. It is found that the variations of JPEG file sizes are on average 3.41% and 38.21% for our method and [5], respectively. Therefore, in terms of image quality and file size variation, our proposed scheme is superior to that method in [5], with the aid of SA-based optimization. Beside, NQMs of the proposed scheme are higher than those of [5] except the case of L=8192, which is beyond real application due to low image quality.

Here we also evaluate the anti-steganalysis performance of the proposed scheme and [5]. The steganalytic scheme in [3] is used for evaluation. A measure, pass rate, is defined as the ratio of the number of stego images not detected by [3] to the number of total test images (i.e., 50). In this experiment, the embedding capacity of the stego images generated by two algorithms is kept the same for fair comparison. From Table I, it can be observed that our scheme has successfully enhanced the anti-steganalysis capability for the steganographic scheme proposed in [5], as the embedding capacity is higher, e.g., at L=2048.

6. CONCLUSIONS

Here, we have demonstrated the effectiveness of our proposed closed-loop system in improving image quality and enhancing the anti-steganalysis capability after data embedding. In principle, this closed-loop architecture can be applied to most types of steganographic schemes (in either spatial or transform domain) and steganalytic systems (either ours in [3] or others published elsewhere). Moreover, more constraints or performance index can be added into the cost function for optimization. Without loss of generality, this architecture is useful in enhancing steganographic schemes in terms of multi-functionalities such as embedding capacity, picture quality, HVS, anti-steganalysis, BER, or more.

ACKNOWLEDGEMENTS

This work was supported in part by National Science Council and Ministry of Economic Affairs, Taiwan, ROC., under the grant number NSC 95-2221-E-212-034 and 95-EC-17-A-02-S1-024, respectively.

REFERENCES

[1] F. A. P. Petitcolas, R. J. Anderson, and M. G. Kuhn, "Information hiding – a survey," *Proceedings of the IEEE*, vol. 87, no. 7, pp. 1062–1078, 1999.

[2] I. Avcibas, N. Memon, and B. Sankur, "Steganalysis using image quality metrics," *IEEE Trans. Image Process.*, vol. 12, no. 2, pp. 221–229, Feb. 2003.

[3] W.-N. Lie and G.-S. Lin, "A feature-based classification technique for blind image steganalysis," *IEEE Trans. Multimedia*, vol.7, no.6, pp.1007-1020, Dec. 2005.

[4] Wen-Nung Lie, Guo-Shiang Lin, and Shen-Long Cheng, "Dual protection of JPEG images based on informed embedding and two-stage watermark extraction," *IEEE Trans. Information Forensics and Security*, vol. 1, no. 3, pp. 330-341, Sep. 2006.

[5] Y. Seki, H. Kobayashi, M. Fujiyoshi, and H. Kiya, "Quantization-based image steganography without data hiding position memorization," *Proc. IEEE Int'l Symposium on Circuits and Systems*, vol. 5, pp. 4987–4990, 2005.

[6] D. J. Ram, T. H. Sreenivas, and K. G. Subramaniam, "Parallel simulated annealing algorithms," *Journal of Parallel and Distributed Computing*, vol. 37, pp. 207–212, 1996.

[7] M. L. Miller, G. J. Doerr, and I. J. Cox, "Applying informed coding and embedding to design a robust high-capacity watermark," *IEEE Trans. Image Processing*, vol. 13, no. 6, pp. 792–807, 2004.

[8] D. Levicky and P. Foris, "Human visual system models in digital image watermarking**,"** *Radioengineering*, vol. 13, no. 4, pp. 38–43, Dec. 2004.

[9] C.-Y. Lin and S.-F. Chang, "Watermarking capacity of digital images based on domain-specific masking effects," *Proc. Int'l Conf. on Information Technology: Coding and Computing*, pp. 90-94, 2001.

[10] N. Damera-Venkata, T.-D. Kite, W.-S. Geisler, B.-L. Evans, and A.- C. Bovik, "Image quality assessment based on a degradation model," *IEEE Trans. on Image Processing,* vol. 9, no. 4, pp. 636-650, 2000.

[11] Y.-T. Wu and F. Y. Shih, "Genetic algorithm based methodology for breaking the steganalytic systems," *IEEE Trans. on Systems, Man, and Cybernetics-Part B*, vol. 36, no. 1, pp. 25-31, 2006.

[12] K. Solanki, K. Sullivan, U. Madhow, B.S. Manjunath, and S. Chandrasekaran, "Statistical restoration for robust and secure steganography," *IEEE Int'l Conf. on Image Process.*, Vol.2, pp.11-14, 2005.

Table I PSNR, file size variation, and NQM of the proposed scheme, compared to [5].

L	PSNR (dB)		Variation in file size		NQM (dB) [10]		Pass rate	
	Proposed	[5]	Proposed	[5]	Proposed	[5]	Proposed	[5]
1024	42.13	36.33	0.29%	6.11%	41.06	37.76	100%	100%
2048	38.24	30.57	0.87%	18.87%	38.16	33.21	100%	96%
4096	31.94	24.67	3.47%	43.86%	30.26	29.35	100%	6%
8192	27.00	21.79	9.01%	84.02%	24.07	24.32	90%	0%