mean and the variance $\alpha^2\sigma_x^2$, and let μ denote the Gaussian distribution with zero mean and the variance σ_x^2. Then, the function f^* is given as follows:

$$\mathrm{f}^*(t) = P_\nu^{-1}\left(P_\mu(t)\right) \tag{31}$$

where P_ν^{-1} is the inverse cumulative distribution function of the Gaussian distribution ν, and P_μ is the cumulative distribution function of the Gaussian distribution μ. It is easy to show that

$$P_\nu^{-1}(z) = \sqrt{2}\alpha\sigma_x \,\mathrm{erf}^{-1}(2z-1), \tag{32}$$

$$P_\mu(t) = \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{t}{\sigma_x\sqrt{2}}\right)\right). \tag{33}$$

By substituting (32) and (33) into (31), we can deduce that

$$\mathrm{f}^*(t) = \alpha t \tag{34}$$

which implies that the optimal way of altering the host signal on $\mathbf{v}_j$ is given by $s_v(j) = \alpha x_v(j)$.

REFERENCES

[1] P. Comesana, L. Perez-Freire, and F. Perez-Gonzalez, "Fundamentals of data hiding security and their application to spread-spectrum analysis," in *Proc. 7th Inf. Hiding Workshop (IH 2005), Lectures Notes in Computer Science, Springer-Verlag*, Barcelona, Spain, Jun. 2005, vol. 3727, pp. 146–160.

[2] F. Cayre, C. Fontaine, and T. Furon, "Watermarking security: Theory and practice," *IEEE Trans. Signal Process.*, vol. 53, no. 10, pp. 3976–3987, Oct. 2005.

[3] I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Trans. Image Process.*, vol. 6, no. 12, pp. 1673–1687, Dec. 1997.

[4] P. Moulin and A. Ivanovic, "The zero-rate spread-spectrum watermarking game," *IEEE Trans. Signal Process.*, vol. 51, no. 4, pp. 1098–1117, Apr. 2003.

[5] H. S. Malvar and D. A. F. Florencio, "Improved spread spectrum: A new modulation technique for robust watermarking," *IEEE Trans. Signal Process.*, vol. 51, no. 4, pp. 898–905, Apr. 2003.

[6] P. Bas and F. Cayre, "Natural watermarking: A secure spread spectrum technique for woa," in *Proc. 8th Inf. Hiding Workshop (IH 2006), Lecture Notes in Computer Science, Springer-Verlag*, Alexandria, VA, Jul. 2006, vol. 4437, pp. 1–14.

[7] B. Mathon, P. Bas, F. Cayre, and B. Macq, "Optimization of natural watermarking using transportation theory," in *Proc. 11th ACM Workshop on Multimedia and Security (MM&Sec'2009)*, Princeton, NJ, Sep. 2009, pp. 33–38.

[8] P. Bas and F. Cayre, "Achieving subspace or key security for woa using natural or circular watermarking," in *Proc. 8th ACM Workshop Multimedia and Security (MM&Sec'06)*, Geneva, Switzerland, Sep. 2006, pp. 80–88.

[9] F. Cayre and P. Bas, "Kerckhoffs-based embedding security classes for woa data hiding," *IEEE Trans. Inf. Forensics Security*, vol. 3, no. 1, pp. 1–15, Mar. 2008.

[10] J. Cao, J. Huang, and J. Ni, "A new spread-spectrum watermarking scheme to achieve a trade-off between security and robustness," in *Proc. 12th Inf. Hiding Workshop (IH 2010), Lectures Notes in Computer Science, Springer-Verlag*, Calgary, AB, Canada, Jun. 2010, vol. 6387, pp. 262–276.

[11] L. V. Kantorovich, "On a problem of Monge," *Uspekhi Math. Nauk.*, vol. 3, pp. 225–226, 1948.

[12] M. Knott and C. S. Smith, "On the optimal mapping of distributions," *J. Optimiz. Theory Applicat.*, vol. 43, pp. 39–49, 1984.

[13] L. V. Kantorovich, "On the translocation of masses," *C. R. (Doklady) Acad. Sci. URSS (N.S.)*, vol. 37, pp. 199–201, 1942.

Separable Reversible Data Hiding in Encrypted Image

Xinpeng Zhang

***Abstract*—This work proposes a novel scheme for separable reversible data hiding in encrypted images. In the first phase, a content owner encrypts the original uncompressed image using an encryption key. Then, a data-hider may compress the least significant bits of the encrypted image using a data-hiding key to create a sparse space to accommodate some additional data. With an encrypted image containing additional data, if a receiver has the data-hiding key, he can extract the additional data though he does not know the image content. If the receiver has the encryption key, he can decrypt the received data to obtain an image similar to the original one, but cannot extract the additional data. If the receiver has both the data-hiding key and the encryption key, he can extract the additional data and recover the original content without any error by exploiting the spatial correlation in natural image when the amount of additional data is not too large.**

***Index Terms*—Image encryption, image recovery, reversible data hiding.**

I. INTRODUCTION

In recent years, signal processing in the encrypted domain has attracted considerable research interest. As an effective and popular means for privacy protection, encryption converts the ordinary signal into unintelligible data, so that the traditional signal processing usually takes place before encryption or after decryption. However, in some scenarios that a content owner does not trust the processing service provider, the ability to manipulate the encrypted data when keeping the plain content unrevealed is desired. For instance, when the secret data to be transmitted are encrypted, a channel provider without any knowledge of the cryptographic key may tend to compress the encrypted data due to the limited channel resource. While an encrypted binary image can be compressed with a lossless manner by finding the syndromes of low-density parity-check codes [1], a lossless compression method for encrypted gray image using progressive decomposition and rate-compatible punctured turbo codes is developed in [2]. With the lossy compression method presented in [3], an encrypted gray image can be efficiently compressed by discarding the excessively rough and fine information of coefficients generated from orthogonal transform. When having the compressed data, a receiver may reconstruct the principal content of original image by retrieving the values of coefficients. The computation of transform in the encrypted domain has also been studied. Based on the homomorphic properties of the underlying cryptosystem, the discrete Fourier transform in the encrypted domain can be implemented [4]. In [5], a composite signal representation method packing together a number of signal samples and processing them as a unique sample is used to reduce the complexity of computation and the size of encrypted data.

Manuscript received June 20, 2011; revised November 07, 2011; accepted November 08, 2011. Date of publication November 15, 2011; date of current version March 08, 2012. This work was supported by the National Natural Science Foundation of China under Grant 61073190, Grant 61103181, and Grant 60832010, by the Research Fund for the Doctoral Program of Higher Education of China under Grant 20113108110010, and by the Alexander von Humboldt Foundation. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Alessandro Piva.

The author is with the School of Communication, Shanghai University, Shanghai 200072, China (e-mail: xzhang@shu.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIFS.2011.2176120

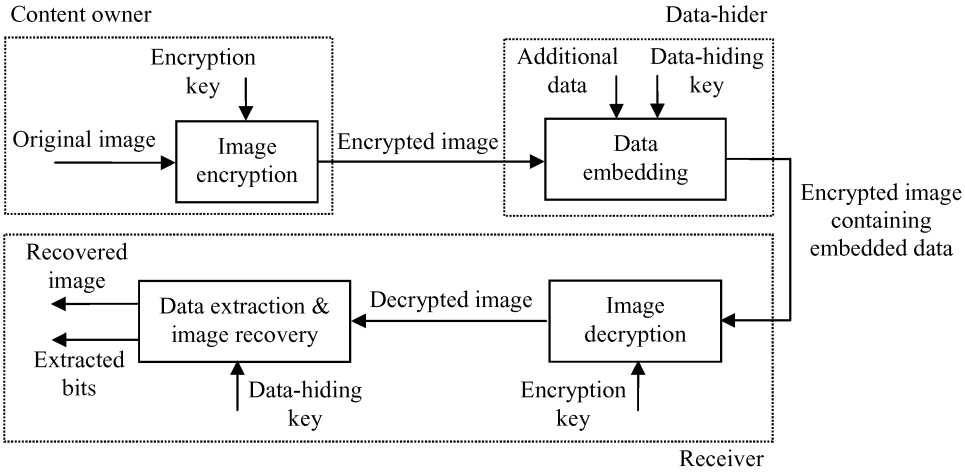

Fig. 1. Sketch of non-separable reversible data hiding in encrypted image.

There are also a number of works on data hiding in the encrypted domain. In a buyer–seller watermarking protocol [6], the seller of digital multimedia product encrypts the original data using a public key, and then permutes and embeds an encrypted fingerprint provided by the buyer in the encrypted domain. After decryption with a private key, the buyer can obtain a watermarked product. This protocol ensures that the seller cannot know the buyer's watermarked version while the buyer cannot know the original version. An anonymous fingerprinting scheme that improves the enciphering rate by exploiting the Okamoto-Uchiyama encryption method has been proposed in [7]. By introducing the composite signal representation mechanism, both the computational overhead and the large communication bandwidth due to the homomorphic public-key encryption are also significantly reduced [8]. In another type of joint data-hiding and encryption schemes, a part of cover data is used to carry the additional message and the rest of the data are encrypted, so that both the copyright and the privacy can be protected. For example [9], the intraprediction mode, motion vector difference and signs of DCT coefficients are encrypted, while a watermark is embedded into the amplitudes of DCT coefficients. In [10], the cover data in higher and lower bit-planes of transform domain are respectively encrypted and watermarked. In [11], the content owner encrypts the signs of host DCT coefficients and each content-user uses a different key to decrypt only a subset of the coefficients, so that a series of versions containing different fingerprints are generated for the users.

The reversible data hiding in encrypted image is investigated in [12]. Most of the work on reversible data hiding focuses on the data embedding/extracting on the plain spatial domain [13]–[17]. But, in some applications, an inferior assistant or a channel administrator hopes to append some additional message, such as the origin information, image notation or authentication data, within the encrypted image though he does not know the original image content. And it is also hopeful that the original content should be recovered without any error after image decryption and message extraction at receiver side. Reference [12] presents a practical scheme satisfying the above-mentioned requirements and Fig. 1 gives the sketch. A content owner encrypts the original image using an encryption key, and a data-hider can embed additional data into the encrypted image using a data-hiding key though he does not know the original content. With an encrypted image containing additional data, a receiver may first decrypt it according to the encryption key, and then extract the embedded data and recover the original image according to the data-hiding key. In the scheme, the data extraction is not separable from the content decryption. In other words, the additional data must be extracted from the decrypted image, so that the principal content of original image is revealed before data extraction, and, if someone has the data-hiding key but not the encryption key, he cannot extract any information from the encrypted image containing additional data.

This paper proposes a novel scheme for separable reversible data hiding in encrypted image. In the proposed scheme, the original image is encrypted using an encryption key and the additional data are embedded into the encrypted image using a data-hiding key. With an encrypted image containing additional data, if the receiver has only the data-hiding key, he can extract the additional data though he does not know the image content. If he has only the encryption key, he can decrypt the received data to obtain an image similar to the original one, but cannot extract the embedded additional data. If the receiver has both the data-hiding key and the encryption key, he can extract the additional data and recover the original image without any error when the amount of additional data is not too large.

II. Proposed Scheme

The proposed scheme is made up of image encryption, data embedding and data-extraction/image-recovery phases. The content owner encrypts the original uncompressed image using an encryption key to produce an encrypted image. Then, the data-hider compresses the least significant bits (LSB) of the encrypted image using a data-hiding key to create a sparse space to accommodate the additional data. At the receiver side, the data embedded in the created space can be easily retrieved from the encrypted image containing additional data according to the data-hiding key. Since the data embedding only affects the LSB, a decryption with the encryption key can result in an image similar to the original version. When using both of the encryption and data-hiding keys, the embedded additional data can be successfully extracted and the original image can be perfectly recovered by exploiting the spatial correlation in natural image. Fig. 2 shows the three cases at the receiver side.

A. Image Encryption

Assume the original image with a size of $N_1 \times N_2$ is in uncompressed format and each pixel with gray value falling into [0, 255] is represented by 8 bits. Denote the bits of a pixel as $b_{i,j,0}, b_{i,j,1}, \ldots, b_{i,j,7}$ where $1 \le i \le N_1$ and $1 \le j \le N_2$, the gray value as $p_{i,j}$, and the number of pixels as $N (N = N_1 \times N_2)$. That implies

$$b_{i,j,u} = \lfloor p_{i,j}/2^u \rfloor \bmod 2, \quad u = 0, 1, \ldots, 7 \tag{1}$$

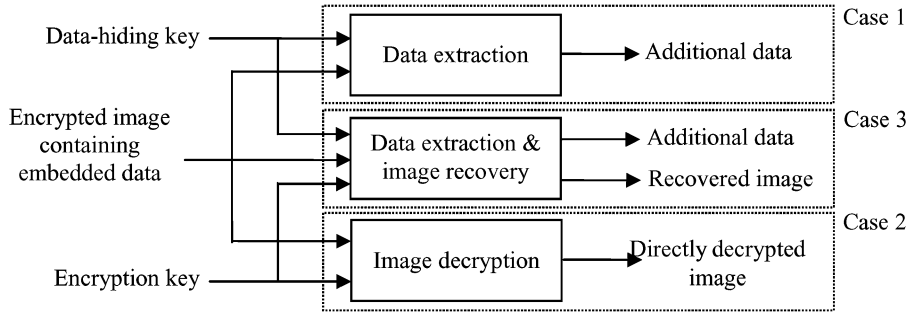

Fig. 2. Three cases at receiver side of the proposed separable scheme.

TABLE I
THEORETICAL VALUES OF PSNR (DB) WITH RESPECT TO S AND M

	$S=1$	$S=2$	$S=3$	$S=4$	$S=5$
$M=1$	54.2	52.4	51.7	51.4	51.3
$M=2$	47.2	45.4	44.7	44.4	44.3
$M=3$	40.9	39.1	38.5	38.2	38.1

and

$$p_{i,j} = \sum_{u=0}^{7} b_{i,j,u} \cdot 2^u. \tag{2}$$

In encryption phase, the exclusive-or results of the original bits and pseudo-random bits are calculated

$$B_{i,j,u} = b_{i,j,u} \oplus r_{i,j,u} \tag{3}$$

where $r_{i,j,u}$ are determined by an encryption key using a standard stream cipher. Then, $B_{i,j,u}$ are concatenated orderly as the encrypted data.

B. Data Embedding

In the data embedding phase, some parameters are embedded into a small number of encrypted pixels, and the LSB of the other encrypted pixels are compressed to create a space for accommodating the additional data and the original data at the positions occupied by the parameters. The detailed procedure is as follows.

According to a data-hiding key, the data-hider pseudo-randomly selects N_{P} encrypted pixels that will be used to carry the parameters for data hiding. Here, N_{P} is a small positive integer, for example, $N_{\mathrm{P}} = 20$. The other $(N - N_{\mathrm{P}})$ encrypted pixels are pseudo-randomly permuted and divided into a number of groups, each of which contains L pixels. The permutation way is also determined by the data-hiding key. For each pixel-group, collect the M least significant bits of the L pixels, and denote them as $B(k,1), B(k,2), \ldots, B(k, M \cdot L)$ where k is a group index within $[1, (N - N_{\mathrm{P}})/L]$ and M is a positive integer less than 5. The data-hider also generates a matrix $\mathbf{G}$ sized $(M \cdot L - S) \times M \cdot L$, which is composed of two parts

$$\mathbf{G} = [\mathbf{I}_{M \cdot L - S}\, \mathbf{Q}]. \tag{4}$$

While the left part is an $(M \cdot L - S) \times (M \cdot L - S)$ identity matrix, the right part $\mathbf{Q}$ sized $(M \cdot L - S) \times S$ is a pseudo-random binary matrix derived from the data-hiding key. Here, S is a small positive integer. Then, embed the values of the parameters M, L and S into the LSB of N_{P} selected encrypted pixels. For the example of $N_{\mathrm{P}} = 20$, the data-hider may represent the values of M, L and S as 2, 14 and 4 bits, respectively, and replace the LSB of selected encrypted pixels with the 20 bits.

In the following, a total of $(N - N_{\mathrm{P}}) \cdot S/L$ bits made up of N_{P} original LSB of selected encrypted pixels and $(N - N_{\mathrm{P}}) \cdot S/L - N_{\mathrm{P}}$ additional bits will be embedded into the pixel groups. For each group, calculate

$$\begin{bmatrix} B'(k,1) \\ B'(k,2) \\ \vdots \\ B'(k, ML - S) \end{bmatrix} = \mathbf{G} \cdot \begin{bmatrix} B(k,1) \\ B(k,2) \\ \vdots \\ B(k, ML) \end{bmatrix} \tag{5}$$

where the arithmetic is modulo-2. By (5), $[B(k,1), B(k,2), \ldots, B(k, M \cdot L)]$ are compressed as $(M \cdot L - S)$ bits, and a sparse space is therefore available for data accommodation. Let $[B'(k, M \cdot L - S + 1), B'(k, M \cdot L - S + 2), \ldots, B'(k, M \cdot L)]$ of each group be the original LSB of selected encrypted pixels and the additional data to be embedded. Then, replace the $[B(k,1), B(k,2), \ldots, B(k, M \cdot L)]$ with the new $[B'(k,1), B'(k,2), \ldots, B'(k, M \cdot L)]$, and put them into their original positions by an inverse permutation. At the same time, the (8-M) most significant bits (MSB) of encrypted pixels are kept unchanged. Since S bits are embedded into each pixel-group, the total $(N - N_{\mathrm{p}}) \cdot S/L$ bits can be accommodated in all groups. Clearly, the embedding rate, a ratio between the data amount of net payload and the total number of cover pixels, is

$$R = \frac{((N - N_P) \cdot S/L - N_P)}{N} \approx \frac{S}{L}. \tag{6}$$

C. Data Extraction and Image Recovery

In this phase, we will consider the three cases that a receiver has only the data-hiding key, only the encryption key, and both the data-hiding and encryption keys, respectively.

With an encrypted image containing embedded data, if the receiver has only the data-hiding key, he may first obtain the values of the parameters M, L and S from the LSB of the N_{P} selected encrypted pixels. Then, the receiver permutes and divides the other $(N - N_{\mathrm{P}})$ pixels into $(N - N_{\mathrm{P}})/L$ groups and extracts the S embedded bits from the M LSB-planes of each group. When having the total $(N - N_{\mathrm{P}}) \cdot S/L$ extracted bits, the receiver can divide them into N_{P} original LSB of selected encrypted pixels and $(N - N_{\mathrm{P}}) \cdot S/L - N_{\mathrm{P}}$ additional bits. Note that because of the pseudo-random pixel selection and permutation, any attacker without the data-hiding key cannot obtain the parameter values and the pixel-groups, therefore cannot extract the embedded data. Furthermore, although the receiver having the data-hiding key can

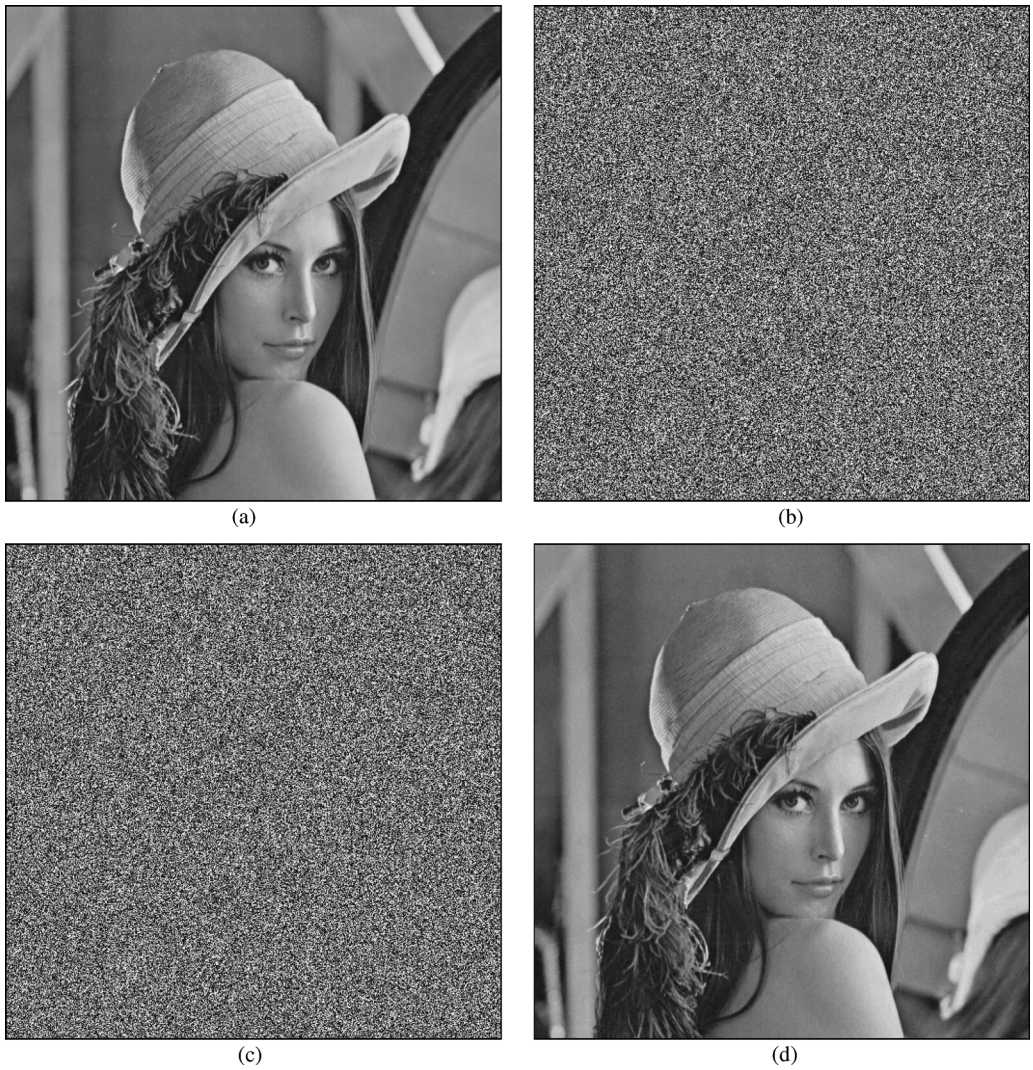

Fig. 3. (a) Original Lena, (b) its encrypted version, (c) encrypted image containing embedded data with embedding rate 0.017 bpp, and (d) directly decrypted version with PSNR 39.0 dB.

successfully extract the embedded data, he cannot get any information about the original image content.

Consider the case that the receiver has the encryption key but does not know the data-hiding key. Clearly, he cannot obtain the values of parameters and cannot extract the embedded data. However, the original image content can be roughly recovered. Denoting the bits of pixels in the encrypted image containing embedded data as $B'_{i,j,0}, B'_{i,j,1}, \ldots, B'_{i,j,7}$ ($1 \le i \le N_1$ and $1 \le j \le N_2$), the receiver can decrypt the received data

$$b'_{i,j,u} = B'_{i,j,u} \oplus r_{i,j,u} \tag{7}$$

where $r_{i,j,u}$ are derived from the encryption key. The gray values of decrypted pixels are

$$p'_{i,j} = \sum_{u=0}^{7} b'_{i,j,u} \cdot 2^u. \tag{8}$$

Since the data-embedding operation does not alter any MSB of encrypted image, the decrypted MSB must be same as the original MSB. So, the content of decrypted image is similar to that of original image. According to (5), if $B(k, M \cdot L - S + 1) = B(k, M \cdot L - S + 2) = \ldots = B(k, M \cdot L) = 0$, there is

$$B'(k, v) = B(k, v), \quad v = 1, 2, \ldots, ML - S. \tag{9}$$

The probability of this case is $1/2^S$, and, in this case, the original $(M \cdot L - S)$ bits in the M LSB-planes can be correctly decrypted. Since S is significantly less than $M \cdot L$, we ignore the distortion at other S decrypted bits. If there are nonzero bits among $B(k, M \cdot L - S + 1)$, $B(k, M \cdot L - S + 2), \ldots$, and $B(k, M \cdot L)$, the encrypted data in the M LSB-planes have been changed by the data-embedding operation, so that the decrypted data in the M LSB-planes differ from the original data. Assuming that the original distribution of the data in the M LSB-planes is uniform, the distortion energy per each decrypted pixel is

$$D_{\mathrm{E}} = 2^{-2M} \cdot \sum_{\alpha=0}^{2^M-1} \sum_{\beta=0}^{2^M-1} (\alpha - \beta)^2. \tag{10}$$

Because the probability of this case is $(2^S-1)/2^S$, the average energy of distortion is

$$A_{\rm E} = \frac{(2^S-1)}{2^S}\cdot 2^{-2M}\cdot \sum_{\alpha=0}^{2^M-1}\sum_{\beta=0}^{2^M-1}(\alpha-\beta)^2. \quad (11)$$

Here, the distortion in the $N_{\rm P}$ selected pixels is also ignored since their number is significantly less than the image size N. So, the value of PSNR in the directly decrypted image is

$$\mathrm{PSNR} = 10\cdot\log_{10}(A_{\rm E}). \quad (12)$$

Table I gives the theoretical values of PSNR with respect to S and M.

If the receiver has both the data-hiding and the encryption keys, he may aim to extract the embedded data and recover the original image. According to the data-hiding key, the values of M, L and S, the original LSB of the $N_{\rm P}$ selected encrypted pixels, and the $(N-N_{\rm P})\cdot S/L - N_{\rm P}$ additional bits can be extracted from the encrypted image containing embedded data. By putting the $N_{\rm P}$ LSB into their original positions, the encrypted data of the $N_{\rm P}$ selected pixels are retrieved, and their original gray values can be correctly decrypted using the encryption keys. In the following, we will recover the original gray values of the other $(N-N_{\rm P})$ pixels. Considering a pixel-group, because $B'(k,1), B'(k2),\ldots,B'(k,M\cdot L-S)$ in (5) are given, $[B(k,1),B(k,2),\ldots,B(k,M\cdot L)]^T$ must be one of the vectors meeting

$$\mathbf{v} = \left[B'(k,1)B'(k,2)\cdots B'(k,ML-S)00\cdots 0\right]^T + \mathbf{a}\cdot\mathbf{H} \quad (13)$$

where $\mathbf{a}$ is an arbitrary binary vector sized $1\times S$, and $\mathbf{H}$ is an $S\times ML$ matrix made up of the transpose of $\mathbf{Q}$ and an $S\times S$ identity matrix

$$\mathbf{H} = [\mathbf{Q}^T\ \mathbf{I}_S]. \quad (14)$$

In other words, with the constraint of (5), there are 2^S possible solutions of $[B(k,1),B(k,2),\ldots,B(k,M\cdot L)]^{\rm T}$. For each vector $\mathbf{v}$, we attempt to put the elements in it to the original positions to get an encrypted pixel-group and then decrypt the pixel-group using the encryption key. Denoting the decrypted pixel-group as G_k and the gray values in it as $t_{i,j}$, calculate the total difference between the decrypted and estimated gray values in the group

$$D = \sum_{(i,j)\in G_k}|t_{i,j}-\tilde{p}_{i,j}| \quad (15)$$

where the estimated gray values is generated from the neighbors in the directly decrypted image, by (16), as shown at the bottom of the page. Clearly, the estimated gray values in (16) are only dependent on the MSB of neighbor pixels. Thus, we have 2^S different D corresponding to the 2^S decrypted pixel-group G_k. Among the 2^S decrypted pixel-group, there must be one that is just the original gray values and possesses a low D because of the spatial correlation in natural image. So, we find the smallest D and regard the corresponding vector $\mathbf{v}$ as the actual $[B(k,1),B(k,2),\ldots,B(k,M\cdot L)]^{\rm T}$ and the decrypted $t_{i,j}$ as the recovered content. As long as the number of pixels in a group is sufficiently large and there are not too many bits embedded into each group, the original content can be perfectly recovered by the spatial correlation criterion. Since the 2^S different D must be calculated in each group, the computation complexity of the content recovery is $O(N\cdot 2^S)$. On the other hand, if more neighboring pixels and a smarter prediction method are used to estimate the gray values, the performance of content recovery will be better, but the computation complexity is higher. To keep a low computation complexity, we let S be less than ten and use only the four neighboring pixels to calculate the estimated values as in (16).

III. Experimental Results

The test image Lena sized 512×512 shown in Fig. 3(a) was used as the original image in the experiment. After image encryption, the eight encrypted bits of each pixel are converted into a gray value to generate an encrypted image shown in Fig. 3(b). Then, we let $M=3$, $L=128$ and $S=2$ to embed 4.4×10^3 additional bits into the encrypted image. The encrypted image containing the embedded data is shown in Fig. 3(c), and the embedding rate R is 0.017 bit per pixel (bpp). With an encrypted image containing embedded data, we could extract the additional data using the data-hiding key. If we directly decrypted the encrypted image containing embedded data using the encryption key, the value of PSNR in the decrypted image was 39.0 dB, which verifies the theoretical value 39.1 dB calculated by (12). The directly decrypted image is given as Fig. 3(d). By using both the data-hiding and the encryption keys, the embedded data could be successfully extracted and the original image could be perfectly recovered from the encrypted image containing embedded data.

Tables II and III list the embedding rates, PSNR in directly decrypted images and PSNR in recovered images when different M, L and S were used for images Lena and Man. As analyzed in (6), the embedding rate is dependent on S and L, and the larger S and the smaller L correspond to a higher embedding rate. On the other hand, the smaller the values of M and S, the quality of directly decrypted image is better since more data in encrypted image are not changed by data embedding. The "$+\infty$" in Tables II and III indicate that the original images were recovered without any error. Here, the large M, L and the small S are helpful to the perfect content recovery since more cover data and less possible solutions are involved in the recovery procedure. If M and L are too small or S is too large, the recovery of original content may be unsuccessful, and the values of PSNR in recovered images are also given in Tables II and III.

Figs. 4–6 show the rate-distortion curves of the four images Lena, Man, Lake and Baboon. Here, three quality metrics were used to measure the distortion in directly decrypted image: PSNR, the Watson metric and a universal quality index Q. While PSNR simply indicates the energy of distortion caused by data hiding, the Watson metric is designed by using characteristics of the human visual system and measures the total perceptual error, which is DCT-based and takes into account three factors: contrast sensitivity, luminance masking and contrast masking [18]. Additionally, the quality index Q works in

$$\tilde{p}_{i,j} = \frac{\lfloor p'_{i-1,j}/2^M\rfloor + \lfloor p'_{i+1,j}/2^M\rfloor + \lfloor p'_{i,j-1}/2^M\rfloor + \lfloor p'_{i,j+1}/2^M\rfloor}{4}\cdot 2^M + 2^{M-1} \quad (16)$$

TABLE II
EMBEDDING RATE R, PSNR IN DIRECTLY DECRYPTED IMAGES (DB) AND PSNR IN RECOVERED IMAGES (DB) WITH DIFFERENT PARAMETERS FOR TEST IMAGE LENA

		$S=1$	$S=2$	$S=3$	$S=4$	$S=5$
$M=1$	$L=2000$	0.0005, 54.6, $+\infty$	0.0010, 52.3, $+\infty$	0.0015, 51.6, $+\infty$	0.0020, 51.4, $+\infty$	0.0025, 51.3, $+\infty$
	$L=1500$	0.0067, 54.3, $+\infty$	0.0013, 52.2, $+\infty$	0.0020, 51.7, $+\infty$	0.0027, 51.4, $+\infty$	0.0033, 51.3, 73.8
	$L=1000$	0.0010, 54.3, $+\infty$	0.0020, 52.2, $+\infty$	0.0030, 51.5, 70.4	0.0040, 51.3, 68.2	0.0050, 51.2, 70.6
$M=2$	$L=400$	0.0025, 47.7, $+\infty$	0.0050, 45.3, $+\infty$	0.0075, 44.7, $+\infty$	0.010, 44.3, $+\infty$	0.013, 44.2, 72.6
	$L=300$	0.0033, 47.5, $+\infty$	0.0067, 45.3, $+\infty$	0.010, 44.6, $+\infty$	0.013, 44.4, 73.6	0.017, 44.2, 70.2
	$L=200$	0.005, 47.6, $+\infty$	0.010, 45.2, $+\infty$	0.015, 44.7, 68.3	0.020, 44.4, 62.6	0.025, 44.2, 61.9
$M=3$	$L=150$	0.007, 41.4, $+\infty$	0.013, 39.0, $+\infty$	0.020, 38.4, $+\infty$	0.027, 38.1, $+\infty$	0.033, 38.0, $+\infty$
	$L=125$	0.008, 41.4, $+\infty$	0.016, 39.0, $+\infty$	0.024, 38.5, $+\infty$	0.032, 38.1, $+\infty$	0.040, 38.0, 71.5
	$L=100$	0.010, 41.0, $+\infty$	0.020, 39.0, $+\infty$	0.030, 38.5, 67.8	0.040, 38.1, 67.9	0.050, 38.0, 65.3

TABLE III
EMBEDDING RATE R, PSNR IN DIRECTLY DECRYPTED IMAGES (DB) AND PSNR IN RECOVERED IMAGES (DB) WITH DIFFERENT PARAMETERS FOR TEST IMAGE MAN

		$S=1$	$S=2$	$S=3$	$S=4$	$S=5$
$M=1$	$L=2000$	0.0005, 53.8, $+\infty$	0.0010, 52.2, $+\infty$	0.0015, 51.7, $+\infty$	0.0020, 51.4, $+\infty$	0.0025, 51.4, $+\infty$
	$L=1500$	0.0067, 54.2, $+\infty$	0.0013, 52.1, $+\infty$	0.0020, 51.6, $+\infty$	0.0027, 51.3, $+\infty$	0.0033, 51.2, $+\infty$
	$L=1000$	0.0010, 54.1, $+\infty$	0.0020, 52.2, 75.0	0.0030, 51.7, 75.1	0.0040, 51.4, 72.2	0.0050, 51.2, 68.5
$M=2$	$L=400$	0.0025, 47.2, $+\infty$	0.0050, 45.1, $+\infty$	0.0075, 44.6, $+\infty$	0.010, 44.3, $+\infty$	0.013, 44.2, $+\infty$
	$L=320$	0.0031, 47.3, $+\infty$	0.0063, 45.2, $+\infty$	0.0094, 44.6, 73.7	0.0125, 44.4, 70.2	0.0156, 44.2, 70.3
	$L=250$	0.004, 47.1, $+\infty$	0.008, 45.0, 71.6	0.012, 44.5, 66.6	0.016, 44.3, 64.4	0.020, 44.2, 62.5
$M=3$	$L=200$	0.005, 41.2, $+\infty$	0.010, 39.0, $+\infty$	0.015, 38.2, $+\infty$	0.020, 38.0, $+\infty$	0.025, 37.9, $+\infty$
	$L=150$	0.007, 41.4, $+\infty$	0.013, 38.8, 69.3	0.020, 38.2, $+\infty$	0.027, 38.1, 71.0	0.033, 37.9, 70.5
	$L=120$	0.008, 41.1, $+\infty$	0.017, 39.0, $+\infty$	0.025, 38.4, 66.4	0.033, 38.0, 66.6	0.042, 37.9, 63.0

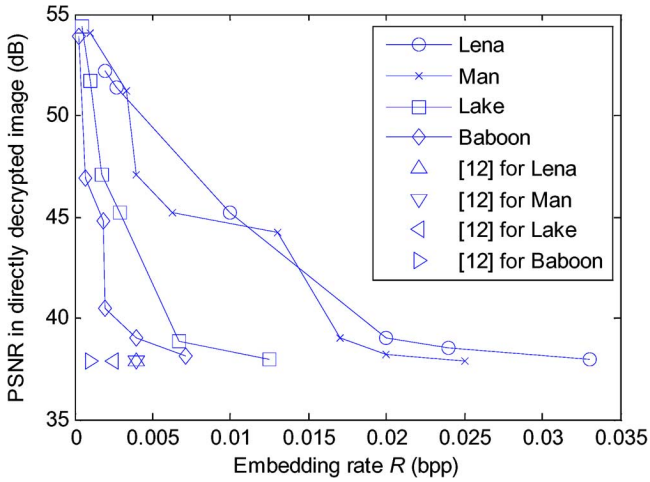

Fig. 4. Rate-PSNR comparison between the proposed scheme and the method in [12].

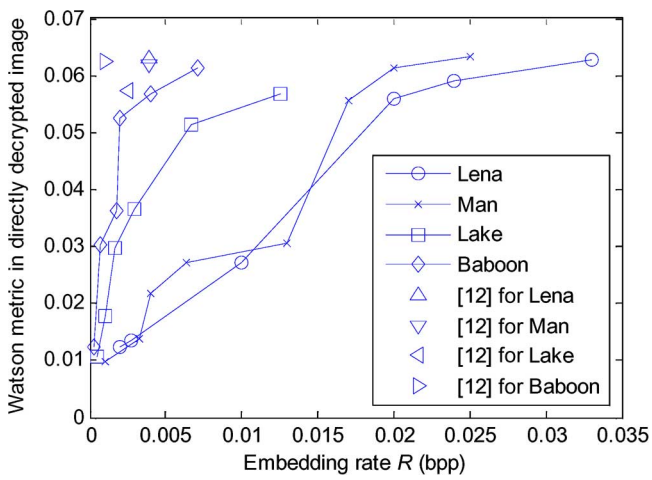

Fig. 5. Rate-Watson metric comparison between the proposed scheme and the method in [12].

spatial domain, as a combination of correlation loss, luminance distortion and contrast distortion [19]. Higher PSNR, lower Watson metric or higher Q means better quality. In these figures, while the abscissa represents the embedding rate, the ordinate is the values of PSNR, Watson metric or quality index Q. The curves are derived from different M, L and S under a condition that the original content can be perfectly recovered using the data-hiding and encryption keys. Since the spatial correlation is exploited for the content recovery, the rate-distortion performance in a smoother image is better. The performance of the nonseparable method in [12] is also given in Figs. 4–6. It can be seen that the performance of the proposed separable scheme is significantly better than that of [12]. We also compared the proposed scheme with the nonseparable method in [12] over 100 images sized 2520 × 3776, which were captured with a digital camera and contain landscape and

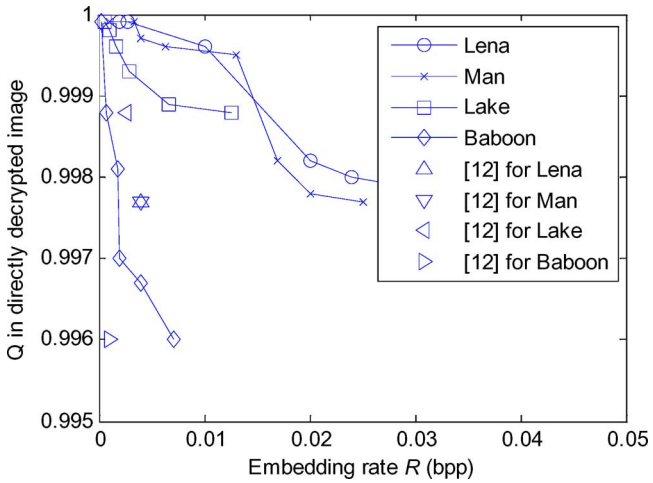

Fig. 6. Rate-Q comparison between the proposed scheme and the method in [12].

people. When meeting the perfect recovery condition, the proposed scheme has an average 203% gain of embedded data amount with same PSNR value in directly decrypted image, or an average gain of 8.7 dB of PSNR value in directly decrypted image with same embedded data amount.

IV. Conclusion

In this paper, a novel scheme for separable reversible data hiding in encrypted image is proposed, which consists of image encryption, data embedding and data-extraction/image-recovery phases. In the first phase, the content owner encrypts the original uncompressed image using an encryption key. Although a data-hider does not know the original content, he can compress the least significant bits of the encrypted image using a data-hiding key to create a sparse space to accommodate the additional data. With an encrypted image containing additional data, the receiver may extract the additional data using only the data-hiding key, or obtain an image similar to the original one using only the encryption key. When the receiver has both of the keys, he can extract the additional data and recover the original content without any error by exploiting the spatial correlation in natural image if the amount of additional data is not too large. If the lossless compression method in [1] or [2] is used for the encrypted image containing embedded data, the additional data can be still extracted and the original content can be also recovered since the lossless compression does not change the content of the encrypted image containing embedded data. However, the lossy compression method in [3] compatible with encrypted images generated by pixel permutation is not suitable here since the encryption is performed by bit-XOR operation. In the future, a comprehensive combination of image encryption and data hiding compatible with lossy compression deserves further investigation.

References

[1] M. Johnson, P. Ishwar, V. M. Prabhakaran, D. Schonberg, and K. Ramchandran, "On compressing encrypted data," *IEEE Trans. Signal Process.*, vol. 52, no. 10, pp. 2992–3006, Oct. 2004.

[2] W. Liu, W. Zeng, L. Dong, and Q. Yao, "Efficient compression of encrypted grayscale images," *IEEE Trans. Image Process.*, vol. 19, no. 4, pp. 1097–1102, Apr. 2010.

[3] X. Zhang, "Lossy compression and iterative reconstruction for encrypted image," *IEEE Trans. Inform. Forensics Security*, vol. 6, no. 1, pp. 53–58, Feb. 2011.

[4] T. Bianchi, A. Piva, and M. Barni, "On the implementation of the discrete Fourier transform in the encrypted domain," *IEEE Trans. Inform. Forensics Security*, vol. 4, no. 1, pp. 86–97, Feb. 2009.

[5] T. Bianchi, A. Piva, and M. Barni, "Composite signal representation for fast and storage-efficient processing of encrypted signals," *IEEE Trans. Inform. Forensics Security*, vol. 5, no. 1, pp. 180–187, Feb. 2010.

[6] N. Memon and P. W. Wong, "A buyer-seller watermarking protocol," *IEEE Trans. Image Process.*, vol. 10, no. 4, pp. 643–649, Apr. 2001.

[7] M. Kuribayashi and H. Tanaka, "Fingerprinting protocol for images based on additive homomorphic property," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2129–2139, Dec. 2005.

[8] M. Deng, T. Bianchi, A. Piva, and B. Preneel, "An efficient buyer-seller watermarking protocol based on composite signal representation," in *Proc. 11th ACM Workshop Multimedia and Security*, 2009, pp. 9–18.

[9] S. Lian, Z. Liu, Z. Ren, and H. Wang, "Commutative encryption and watermarking in video compression," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 6, pp. 774–778, Jun. 2007.

[10] M. Cancellaro, F. Battisti, M. Carli, G. Boato, F. G. B. Natale, and A. Neri, "A commutative digital image watermarking and encryption method in the tree structured Haar transform domain," *Signal Processing: Image Commun.*, vol. 26, no. 1, pp. 1–12, 2011.

[11] D. Kundur and K. Karthik, "Video fingerprinting and encryption principles for digital rights management," *Proceedings IEEE*, vol. 92, no. 6, pp. 918–932, Jun. 2004.

[12] X. Zhang, "Reversible data hiding in encrypted image," *IEEE Signal Process. Lett.*, vol. 18, no. 4, pp. 255–258, Apr. 2011.

[13] J. Tian, "Reversible data embedding using a difference expansion," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 8, pp. 890–896, Aug. 2003.

[14] Z. Ni, Y.-Q. Shi, N. Ansari, and W. Su, "Reversible data hiding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 3, pp. 354–362, Mar. 2006.

[15] M. U. Celik, G. Sharma, A. M. Tekalp, and E. Saber, "Lossless generalized-LSB data embedding," *IEEE Trans. Image Process.*, vol. 14, no. 2, pp. 253–266, Feb. 2005.

[16] W. Hong, T.-S. Chen, Y.-P. Chang, and C.-W. Shiu, "A high capacity reversible data hiding scheme using orthogonal projection and prediction error modification," *Signal Process.*, vol. 90, pp. 2911–2922, 2010.

[17] C.-C. Chang, C.-C. Lin, and Y.-H. Chen, "Reversible data-embedding scheme using differences between original and predicted pixel values," *IET Inform. Security*, vol. 2, no. 2, pp. 35–46, 2008.

[18] A. Mayache, T. Eude, and H. Cherifi, "A comparison of image quality models and metrics based on human visual sensitivity," in *Proc. Int. Conf. Image Processing (ICIP'98)*, Chicago, IL, 1998, vol. 3, pp. 409–413.

[19] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 1, pp. 81–84, Jan. 2002.