PERGAMON

Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT®

Pattern Recognition 37 (2004) 469 – 474

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY

www.elsevier.com/locate/patcog

Hiding data in images by simple LSB substitution

Chi-Kwong Chan*, L.M. Cheng

Department of Computer Engineering and Information Technology, City University of Hong Kong, Hong Kong

Received 17 May 2002; received in revised form 11 July 2003; accepted 11 August 2003

Abstract

In this paper, a data hiding scheme by simple LSB substitution is proposed. By applying an optimal pixel adjustment process to the stego-image obtained by the simple LSB substitution method, the image quality of the stego-image can be greatly improved with low extra computational complexity. The worst case mean-square-error between the stego-image and the cover-image is derived. Experimental results show that the stego-image is visually indistinguishable from the original cover-image. The obtained results also show a significant improvement with respect to a previous work.

Keywords: Data hiding; LSB substitution

1. Introduction

Data hiding is a method of hiding secret messages into a cover-media such that an unintended observer will not be aware of the existence of the hidden messages. In this paper, 8-bit grayscale images are selected as the cover-media. These images are called cover-images. Cover-images with the secret messages embedded in them are called stego-images. For data hiding methods, the image quality refers to the quality of the stego-images.

In the literature, many techniques about data hiding have been proposed [1–5]. One of the common techniques is based on manipulating the least-significant-bit (LSB) planes by directly replacing the LSBs of the cover-image with the message bits. LSB methods typically achieve high capacity.

Wang et al. [6] proposed to embed secret messages in the moderately significant bit of the cover-image. A genetic algorithm is developed to find an optimal substitution matrix for the embedding of the secret messages. They also proposed to use a local pixel adjustment process (LPAP) to improve the image quality of the stego-image. Unfortunately, since the local pixel adjustment process only considers the last three least significant bits and the fourth bit but not on all bits, the local pixel adjustment process is obviously not optimal. The weakness of the local pixel adjustment process is pointed out in Ref. [7]. As the local pixel adjustment process modifies the LSBs, the technique cannot be applied to data hiding schemes based on simple LSB substitution.

Recently, Wang et al. [8] further proposed a data hiding scheme by optimal LSB substitution and genetic algorithm. Using the proposed algorithm, the worst mean-square-error (WMSE) between the cover-image and the stego-image is shown to be $\frac{1}{2}$ of that obtained by the simple LSB substitution method.

In this paper, a data hiding scheme by simple LSB substitution with an optimal pixel adjustment process (OPAP) is proposed. The basic concept of the OPAP is based on the technique proposed in Ref. [7]. The operations of the OPAP is generalized. The WMSE between the cover-image and the stego-image is derived. It is shown that the WMSE obtained by the OPAP could be less than $\frac{1}{2}$ of that obtained by the simple LSB substitution method. Experimental results demonstrate that enhanced image quality can be obtained with low extra computational complexity. The results obtained also show better performance than the optimal substitution method described in Ref. [8].

* Corresponding author.
E-mail address: kwong.chan@alumni.cityu.edu.hk (C.-K. Chan).

0031-3203/$30.00
doi:10.1016/j.patcog.2003.08.007

The rest of the paper is organized as follows. Section 2 briefly describes the simple LSB substitution. In Section 3, the optimal pixel adjustment process is described and the performance is analyzed. Experimental results are given in Section 4. Finally, Section 5 concludes this paper.

2. Data hiding by simple LSB substitution

In this section, the general operations of data hiding by simple LSB substitution method is described.

Let C be the original 8-bit grayscale cover-image of $M_c \times N_c$ pixels represented as

$$C = \{x_{ij} | 0 \leqslant i < M_c, 0 \leqslant j < N_c, \\ x_{ij} \in \{0, 1, \ldots, 255\}\}. \tag{1}$$

M be the n-bit secret message represented as

$$M = \{m_i | 0 \leqslant i < n, m_i \in \{0, 1\}\}. \tag{2}$$

Suppose that the n-bit secret message M is to be embedded into the k-rightmost LSBs of the cover-image C. Firstly, the secret message M is rearranged to form a conceptually k-bit virtual image M' represented as

$$M' = \{m'_i | 0 \leqslant i < n', m'_i \in \{0, 1, \ldots, 2^k - 1\}\}, \tag{3}$$

where $n' < M_c \times N_c$. The mapping between the n-bit secret message $M = \{m_i\}$ and the embedded message $M' = \{m'_i\}$ can be defined as follows:

$$m'_i = \sum_{j=0}^{k-1} m_{i \times k + j} \times 2^{k-1-j}.$$

Secondly, a subset of n' pixels $\{x_{l_1}, x_{l_2}, \ldots, x_{l_{n'}}\}$ is chosen from the cover-image C in a predefined sequence. The embedding process is completed by replacing the k LSBs of x_{l_i} by m'_i. Mathematically, the pixel value x_{l_i} of the chosen pixel for storing the k-bit message m'_i is modified to form the stego-pixel x'_{l_i} as follows:

$$x'_{l_i} = x_{l_i} - x_{l_i} \bmod 2^k + m'_i. \tag{4}$$

In the extraction process, given the stego-image S, the embedded messages can be readily extracted without referring to the original cover-image. Using the same sequence as in the embedding process, the set of pixels $\{x'_{l_1}, x'_{l_2}, \ldots, x'_{l_{n'}}\}$ storing the secret message bits are selected from the stego-image. The k LSBs of the selected pixels are extracted and lined up to reconstruct the secret message bits. Mathematically, the embedded message bits m'_i can be recovered by

$$m'_i = x'_{l_i} \bmod 2^k. \tag{5}$$

Suppose that all the pixels in the cover-image are used for the embedding of secret message by the simple LSB substitution method. Theoretically, in the worst case, the PSNR of the obtained stego-image can be computed by

$$\mathrm{PSNR}_{worst} = 10 \times \log_{10} \frac{255^2}{\mathrm{WMSE}} \\ = 10 \times \log_{10} \frac{255^2}{(2^k - 1)^2} \ \mathrm{dB}. \tag{6}$$

Table 1
Worst PSNR for $k = 1$–5 by simple LSB substitution

k	1	2	3	4	5
PSNR	48.13	38.59	31.23	24.61	18.30

Table 1 tabulates the worst PSNR for some $k = 1$–5. It could be seen that the image quality of the stego-image is degraded drastically when $k \geqslant 4$.

3. Optimal pixel adjustment process

In this section, an optimal pixel adjustment process (OPAP) is proposed to enhance the image quality of the stego-image obtained by the simple LSB substitution method. The basic concept of the OPAP is based on the technique proposed in Ref. [7].

Let p_i, p'_i and p''_i be the corresponding pixel values of the ith pixel in the cover-image C, the stego-image C' obtained by the simple LSB substitution method and the refined stego-image obtained after the OPAP. Let $\delta_i = p'_i - p_i$ be the embedding error between p_i and p'_i. According to the embedding process of the simple LSB substitution method described in Section 2, p'_i is obtained by the direct replacement of the k least significant bits of p_i with k message bits, therefore,

$$-2^k < \delta_i < 2^k. \tag{7}$$

The value of δ_i can be further segmented into three intervals, such that

$$\begin{aligned} &\textit{Interval 1}: && 2^{k-1} < \delta_i < 2^k, \\ &\textit{Interval 2}: && -2^{k-1} \leqslant \delta_i \leqslant 2^{k-1}, \\ &\textit{Interval 3}: && -2^k < \delta_i < -2^{k-1}. \end{aligned} \tag{8}$$

Based on the three intervals, the OPAP, which modifies p'_i to form the stego-pixel p''_i, can be described as follows:

Case 1 ($2^{k-1} < \delta_i < 2^k$): If $p'_i \geqslant 2^k$, then $p''_i = p'_i - 2^k$; otherwise $p''_i = p'_i$;

Case 2 ($-2^{k-1} \leqslant \delta_i \leqslant 2^{k-1}$): $p''_i = p'_i$;

Case 3 ($-2^k < \delta_i < -2^{k-1}$): If $p'_i < 256 - 2^k$, then $p''_i = p'_i + 2^k$; otherwise $p''_i = p'_i$.

Let $\delta_i' = p_i'' - p_i$ be the embedding error between p_i and p_i''. δ_i' can be computed as follows:

Case 1 ($2^{k-1} < \delta_i < 2^k$ and $p_i' \geqslant 2^k$)

$$\begin{aligned}\delta_i' &= p_i'' - p_i = p_i' - 2^k - p_i = \delta_i - 2^k \\ &\Rightarrow 2^{k-1} - 2^k < \delta_i' < 2^k - 2^k \\ &\Rightarrow -2^{k-1} < \delta_i' < 0.\end{aligned}$$

Case 2 ($2^{k-1} < \delta_i < 2^k$ and $p_i' < 2^k$)

$$\begin{aligned}\delta_i' &= p_i'' - p_i = p_i' - p_i = \delta_i \\ &\Rightarrow 2^{k-1} < \delta_i' < 2^k.\end{aligned}$$

Case 3 ($-2^{k-1} \leqslant \delta_i \leqslant 2^{k-1}$)

$$\begin{aligned}\delta_i' &= p_i'' - p_i = p_i' - p_i = \delta_i \\ &\Rightarrow -2^{k-1} \leqslant \delta_i' \leqslant 2^{k-1}.\end{aligned}$$

Case 4 ($-2^k < \delta_i < -2^{k-1}$ and $p_i' < 256 - 2^k$)

$$\begin{aligned}\delta_i' &= p_i'' - p_i = p_i' + 2^k - p_i = \delta_i + 2^k \\ &\Rightarrow -2^k + 2^k < \delta_i' < -2^{k-1} + 2^k \\ &\Rightarrow 0 < \delta_i' < 2^{k-1}.\end{aligned}$$

Case 5 ($-2^k < \delta_i < -2^{k-1}$ and $p_i' \geqslant 256 - 2^k$)

$$\begin{aligned}\delta_i' &= p_i'' - p_i = p_i' - p_i = \delta_i \\ &\Rightarrow -2^k < \delta_i' < -2^{k-1}.\end{aligned}$$

From the above five cases, it can be seen that the absolute value of δ_i' may fall into the range $2^{k-1} < |\delta_i'| < 2^k$ only when $p_i' < 2^k$ (Case 2) and $p_i' \geqslant 256 - 2^k$ (Case 5); while for other possible values of p_i', δ_i' falls into the range $0 \leqslant |\delta_i'| \leqslant 2^{k-1}$. Because p_i' is obtained by the direct replacement of the k LSBs of p_i with the message bits, $p_i' < 2^k$ and $p_i' \geqslant 256 - 2^k$ are equivalent to $p_i < 2^k$ and $p_i \geqslant 256 - 2^k$, respectively. In general, for grayscale natural images, when $k \leqslant 4$, the number of pixels with pixel values smaller than 2^k or greater than $256 - 2^k$ is insignificant. As a result, it could be estimated that the absolute embedding error between pixels in the cover-image and in the stego-image obtained after the proposed OPAP is limited to

boundary pixels

$$\boxed{0 \leqslant |\delta_i'| \leqslant 2^{k-1}.} \tag{9}$$

Let WMSE and WMSE* be the worst case mean-square-error between the stego-image and the cover-image obtained by the simple LSB substitution method and the proposed method with OPAP, respectively. According to Eq. (9) WMSE* can be derived by

$$\mathrm{WMSE}^* = \frac{1}{M_c \times N_c} \sum_{i=0}^{M_c \times N_c - 1} (2^{k-1})^2 = (2^{k-1})^2. \tag{10}$$

Combining Eqs. (6) and (10), we have

$$\boxed{\mathrm{WMSE}^* = \frac{(2^{k-1})^2}{(2^k-1)^2}\,\mathrm{WMSE}}$$

$$= \begin{cases} \mathrm{WMSE} & \text{when } k = 1, \\ \frac{4}{9}\,\mathrm{WMSE} & \text{when } k = 2, \\ \frac{16}{49}\,\mathrm{WMSE} & \text{when } k = 3, \\ \frac{64}{225}\,\mathrm{WMSE} & \text{when } k = 4. \end{cases} \tag{11}$$

Eq. (11) reveals that $\mathrm{WMSE}^* < \frac{1}{2}\,\mathrm{WMSE}$, for $k \geqslant 2$; and $\mathrm{WMSE}^* \approx \frac{1}{4}\,\mathrm{WMSE}$ when $k = 4$. This result also shows that the WMSE* obtained by the OPAP is better than that obtained by the optimal substitution method proposed in Ref. [8] in which $\mathrm{WMSE}^* = \frac{1}{2}\,\mathrm{WMSE}$.

Moreover, the optimal pixel adjustment process only requires a checking of the embedding error between the original cover-image and the stego-image obtained by the simple LSB substitution method to form the final stego-image. The extra computational cost is very small compared with Wang's method [8] which requires huge computation for the genetic algorithm to find an optimal substitution matrix.

4. Experimental results

This section presents experimental results obtained for two cover-image sets. The first set of cover-images consists of four standard grayscale images, 'Lena', 'Baboon', 'Jet' and 'Scene', each of 512 × 512 pixels, as depicted in Fig. 1. The second set consists of 1000 randomly

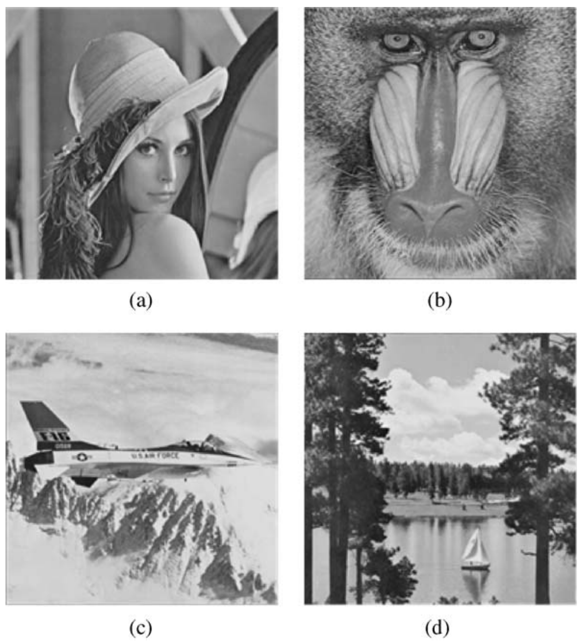

Fig. 1. The first set cover-images of size 512 × 512 pixels.

Fig. 2. Test image used as the second set of secret message.

generated grayscale images. There are two set of secret messages. The first set of secret message consists of 1000 randomly generated message of $512 \times 512 \times k$ bits, where k refers to the number of LSBs in the cover image pixels that are used to hold the secret data bits. For example, suppose that the last two LSBs of the cover image pixels are used to hold the secret data, then the secret data is of size $512 \times 512 \times 2 = 524\,288$ bits. The second set consists of the reduced-sized images of the grayscale image 'Tiff' as shown in Fig. 2. The reduced-sized images are of size 512×256 pixels (for 4-bit insertion), 384×256 pixels (for 3-bit insertion), 256×256 pixels (for 2-bit insertion) and 256×128 pixels (for 1-bit insertion), respectively.

The results of embedding the first set of secret messages into the first set of cover-images are listed in Table 2. Referring to Table 2, the column labeled OPAP is our proposed method with the optimal pixel adjustment process; the column labeled LSB is the simple LSB substitution method; and the column labeled OLSB in the optimal LSB substitution method proposed in Ref. [8]. For the OPAP and LSB methods, the obtained PSNR values are the average values of embedding the 1000 sets random messages into the cover-images. For the OLSB method, for $k = 1, 2$, the obtained PSNR values are the average values of embedding the 1000 sets random messages into the cover-images, for $k = 3$, the obtained PSNR values are the average values of embedding the 10 out of 1000 sets random messages into the cover-images while for $k = 4$, no experiments are conducted due to the large number of searching space for the optimal substitution matrix. The results reveal that our proposed method has much better performance than the LSB and OLSB methods for $k = 2$–4.

The results of embedding the reduced-sized image of Fig. 2 into the first set of cover-images are listed in Table 3. The results also reveal that our proposed method has much better performance than the LSB and OLSB methods for $k = 2$–4.

Table 4 also shows the percentage of cover image pixels associated with the five cases:

$$\begin{aligned}
&\textit{Case } 1 \quad (2^{k-1} < \delta_i < 2^k \text{ and } p_i' \geqslant 2^k),\\
&\textit{Case } 2 \quad (2^{k-1} < \delta_i < 2^k \text{ and } p_i' < 2^k),\\
&\textit{Case } 3 \quad (-2^{k-1} \leqslant \delta_i \leqslant 2^{k-1}),\\
&\textit{Case } 4 \quad (-2^k < \delta_i < -2^{k-1} \text{ and } p_i' < 256 - 2^k),\\
&\textit{Case } 5 \quad (-2^k < \delta_i < -2^{k-1} \text{ and } p_i' \geqslant 256 - 2^k).
\end{aligned} \tag{12}$$

Table 2
The results of embedding the random messages into the first set of cover-images

Cover image	k	OPAP	LSB	OLSB
Lena	1	51.1410	51.1410	51.1483
	2	46.3699	44.1519	44.1651
	3	40.7271	37.9234	37.9467
	4	34.8062	31.7808	—
Baboon	1	51.1414	51.1414	51.1477
	2	46.3691	44.1492	44.1619
	3	40.7253	37.9226	37.9480
	4	34.8021	31.8588	—
Jet	1	51.1405	51.1405	51.1478
	2	46.3700	44.1149	44.1276
	3	40.7273	37.9557	37.9978
	4	34.8065	31.8487	—
Scene	1	51.1410	51.1410	51.1480
	2	46.3702	44.1497	44.1628
	3	40.7270	37.8914	37.9849
	4	34.806	31.8467	—

Table 3
The results of embedding the reduced-sized image of Fig. 2 into the first set of cover-images

Cover image	k	OPAP	LSB	OLSB
Lena	1	51.1299	51.1299	51.1524
	2	46.3707	44.0216	44.7638
	3	40.7266	37.8626	38.7242
	4	34.8434	31.2818	—
Baboon	1	51.1415	51.1415	51.1415
	2	46.3761	44.0205	44.7440
	3	40.7254	37.8642	38.7295
	4	34.7853	31.3307	—
Jet	1	51.1458	51.1458	51.1458
	2	46.3692	43.9901	44.7354
	3	40.7241	37.8898	38.7667
	4	34.8283	31.4083	—
Scene	1	51.1402	51.1402	51.1420
	2	46.3674	44.0176	44.7656
	3	40.7420	37.8522	38.6909
	4	34.7974	31.3208	—

Table 4
The percentage of cover image pixels associated with the five cases (Eq. (12)) when the reduced-sized images of Fig. 2 are embedded into the cover-images

Cover image	k	Case 1 (%)	Case 2 (%)	Case 3 (%)	Case 4 (%)	Case 5
Lena	2	9.52	0	86.55	3.93	0
	3	14.15	0	80.86	4.99	0
	4	21.30	0	73.27	5.43	0
Baboon	2	9.53	0.01	86.51	3.95	0
	3	14.03	0.02	80.90	5.05	0
	4	20.78	0.05	73.85	5.32	0
Jet	2	9.67	0	86.32	4.01	0
	3	13.91	0	81.20	4.89	0
	4	20.31	0	74.22	5.47	0
Scene	2	9.58	0	86.53	3.89	0
	3	14.17	0.01	80.78	5.04	0
	4	21.01	0.01	73.74	5.24	0

when the reduced-sized images of Fig. 2 are embedded into the cover-images.

For illustrative purpose, Fig. 3 shows a pair of stego-images obtained by embedding the reduced-sized image 'Tiff' of size 512×256 pixels into the cover-image 'Lena' of size 512×512 pixels using the simple LSB method and the proposed OPAP method. From Fig. 3(a) (stego-image obtained by the simple LSB-substitution method), one can see some false contours appearing on the shoulder of 'Lena'. The unwanted artifacts may arise suspicion and defeat the purpose of steganography. However, there is no such artifacts appearing on the stego-image (Fig. 3(b)) obtained by the proposed method. The visual quality of stego-images obtained by the proposed method are much better than that of obtained by the simple LSB-substitution method.

To further evaluate the performance of the proposed method, the reduced-sized image of Fig. 2 are embedded into 1000 sets randomly generated cover-images and the obtained average PSNR values are listed in Table 5. The

(a) (b)

Fig. 3. Stego-images obtained by (a) Simple LSB-substitution method; (b) proposed method, where the secret-image is of size 512×256 pixels (4-bit insertion).

Table 5
The results of embedding the reduced-sized image of Fig. 2 into the second set of cover-images

Cover image	k	OPAP	LSB
Random	1	51.1410	51.1410
	2	46.3215	44.0217
	3	40.6023	37.8621
	4	34.4868	31.337

results show that similar PSNR values can be obtained for different type of cover-images.

5. Conclusion

In this paper, a data hiding method by simple LSB substi-
tution with an optimal pixel adjustment process is proposed.
The image quality of the stego-image can be greatly im- 2
1 proved with low extra computational complexity. Extensive
experiments show the effectiveness of the proposed method.
3 The results obtained also show significant improvement than
the method proposed in Ref. [8] with respect to image qual-
ity and computational efficiency.

References

[1] A.Z. Tirkel, R.G. Van Schyndel, C.F. Osborne, A digital watermark, Proceedings of ICIP 1994, Austin Convention Center, Austin, Texas, Vol. II, 1994, pp. 86–90.
[2] W. Bender, N. Morimoto, A. Lu, Techniques for data hiding, IBM Syst. J. 35 (3/4) (1996) 313–336.
[3] T.S. Chen, C.C. Chang, M.S. Hwang, A virtual image cryptosystem based upon vector quantization, IEEE Trans. Image Process. 7 (10) (1998) 1485–1488.
[4] L.M. Marvel, C.G. Boncelet, C.T. Retter, Spread spectrum image steganography, IEEE Trans. Image Process. 8 (8) (1999) 1075–1083.
[5] K.L. Chung, C.H. Shen, L.C. Chang, A novel SVD- and VQ-based image hiding scheme, Pattern Recognition Lett. 22 (9) (2001) 1051–1058.
[6] Ran-Zan Wang, Chi-Fang Lin, Ja-Chen Lin, Hiding data in images by optimal moderately significant-bit replacement, IEE Electron. Lett. 36 (25) (2000) 2069–2070.
[7] Chi-Kwong Chan, L.M. Cheng, Improved hiding data in images by optimal moderately significant-bit replacement, IEE Electron. Lett. 37 (16) (2001) 1017–1018.
[8] Ran-Zan Wang, Chi-Fang Lin, Ja-Chen Lin, Image hiding by optimal LSB substitution and genetic algorithm, Pattern Recognition 34 (3) (2001) 671–683.

About the Author—CHI-KWONG CHAN received the B.Eng. and M.Phil. degrees in Electronic Engineering and the Ph.D. degree in Computer Engineering and Information Technology from City University of Hong Kong in 1996, 1999, and 2003, respectively. His research interests include security, image processing, neural networks, and FPGA implementation.

About the Author—L.M. CHENG received the B.Sc. degree in Physics and Computer Science and the Ph.D. degree from King's College London (previously Queen Elizabeth College), University of London in 1979 and 1982, respectively. He had been employed as a research fellow at King's College, London, Principal Engineer of ERA Technology Ltd. UK, and Project Manager and Senior Consultant at Logica Space and Defence Systems Ltd. UK. He joined City University of Hong Kong in 1989 and is now an Associate Professor at the City University of Hong Kong. His research interests include image processing, security, and neural network.